Vagueness, communication, and semantic information

Sutton, Peter Roger

Awarding institution:
King's College London

Title: Vagueness, communication, and semantic information

Author: Peter Sutton

Vagueness, Communication, and Semantic Information

PhD Thesis in Philosophy

2013

Peter Sutton

King's College London

Primary Supervisor
Professor Shalom Lappin

Secondary Supervisor
Professor Mark Textor

Dedicated to Irene Sutton

"… we speak of people 'taking refuge' in vagueness—the more precise you are, in general the more likely you are to be wrong, whereas you stand a good chance of *not* being wrong if you make it vague enough."

– J. L. Austin, *Sense and Sensibilia*

"The truth is rarely pure and never simple. Modern life would be very tedious if it were either, and modern literature a complete impossibility!"

– Oscar Wilde, *The Importance of Being Earnest*

Contents

Abstract

To be learnable, words must contribute something that is pretty stable across contexts. But equally, words must also be flexible enough to be able to stretch, in a principled way, to cover new cases. Similarly, to be effective for communication, the information that words encode must be robust enough and flexible enough to help us achieve a wide variety of goals. It is argued that truth conditions, and information understood in terms of truth conditions, cannot satisfy these requirements. A replacement for the truth conditional model is suggested based on a statistically grounded conception of semantic information. Informally, this can be understood in terms of reasonable expectations (what it is reasonable to believe, given the words that were used). Formally, this semantic information is captured using probabilistic and information theoretic tools. Vagueness, understood in terms of borderline cases, is argued to be a byproduct of making the above learning and communication requirements central. Vagueness, understood as our ability to be vague with words, is given an information theoretic explanation. Finally, the account is defended with respect to some of the philosophical problems and puzzles found in the vagueness literature.

Acknowledgements

As well as being a huge source of support and guidance, Shalom Lappin has helped me to turn a lot of my sceptical and pessimistic thoughts and energies into what are hopefully now constructive and productive ideas. His flair for being able to speedily perceive and elucidate the major issue in any discussion is breathtaking, and his impact on my philosophical and linguistic outlook is unquantifiable.

Mark Textor's power to identify a hole in an argument and to help fix it, plus his encyclopaedic ability to direct one to a relevant and helpful article is remarkable. A tutorial with Mark results in any piece of work being more informed and greatly improved.

Although not listed officially as one of my supervisors, Ruth Kempson has nonetheless performed that role and much more besides. Her energy, enthusiasm, and perceptiveness are astonishing, as is her ability to view any issue from a novel and insightful perspective. A cup of tea in the sunshine whilst talking with Ruth invariably results either in an exciting new idea to mull over or a novel inspiring way of viewing an old one.

Further to Mark, Ruth, and Shalom, whose impact on this work is immeasurable, I would like to thank the following people for reading and helping me to improve various draft chapters. Alex Clark whose technical prowess, and polite and gentle scepticism have greatly improved the philosophical and formal aspects of this thesis. Mike Coxhead for his frank and elucidating questions and his keen eye for a typo. Wilfried Meyer-Viol, whose sharp formal abilities helped reform and improve almost the entirety of chapter 6. I would also like to thank the following people for insightful and indispensable help with early drafts and papers: Daniel Lassiter, Sarah Murray, Dolf Rami, Peter Ridley, and Will Starr. Many thanks also go to Charles Travis for numerous thought-provoking discussions and inspiring ideas. I am also thankful to the participants of the Bayesian Natural Language Semantics and Pragmatics workshop 2013, the members of the Cornell Semantics Reading Group, the attendees of the Word Meaning Reading Group, and the participants KCL Language and Cognition, and Graduate Seminars for highly useful questions and comments. Thanks also to the members of the DynDial group and others at QMUL for numerous discussions and insights. The responsibility for all errors in this thesis is, nonetheless, entirely mine to bear.

My gratitude goes to the Arts and Humanities Research Council who funded two years of this project. Thanks also to King's College London for the provision of a Sorabji Graduate Bursary, and to both King's College London and Cornell University for the provision of a Norman Malcolm Fellowship. My thanks also go to Dorothy Edgington and Robin Cooper for agreeing to examine this thesis.

I would like to express gratitude to Sven Bernecker, Ahti-Veikko Pietarinen, and Matti Sintonen for their impact on the formative part of my philosophical education. Thanks also to Oskar MacGregor for being an encouragement and an inspiration both whilst we were at the University of Helsinki and thereafter.

I could not have hoped for a more supportive, insightful, and loving group of friends and colleagues at the King's College London Philosophy Department. I would like to express special thanks to all of the following people whom, in more ways than I could mention, have been a source of support, inspiration, help and kindness: Luke Brunning, Lucy Campbell, Mike Campbell, Katherine Jane Cecil, Mike Coxhead, Alex Davies, Helena Drage, Eleni Gregoromichelaki, Sam Kukathas, Chris Machut, Rory O'Connell, Michael O'Sullivan, Peter Ridley, Abby Rutherford, Sarah Tropper, Michael Withey, Jen Wright, and John Wright.

As well as those at King's, a huge thanks must go to my other friends, near and far. In particular, I would not be where I am, nor would I be the person I am without Markus Hippi, Jess Rotas, George Smith and Daniel Snowden. The same applies to all of the *helsinkiläiset*, whom I miss terribly.

My partner David has not only been an amazing tower of strength, support, and love, but has been overwhelmingly understanding of, and throughout, all of the international moves and extended visits that my academic path has thrown up. His patience and tenderness are truly humbling.

My father Roger and my sister Hilary have been remarkably caring, understanding, and loving throughout. The last couple of years would not have been bearable without them.

Finally, there is my mother Irene, who, more than anyone else, gave me a joy of language and a love of books and words. By the time I came back to London in 2009, mum had already been diagnosed with cancer. She died a little over a year later in 2010. I think of her every day. Her quiet strength and resilience are an inspiration in bad times, and her modesty and humanity are an aspiration in good times. I dedicate this thesis to her.

Punctuation and Typeface Conventions

The following will be punctuational conventions throughout:

(i) Double quotes: Direct quotations such as, “truth is supertruth”/Mary said, “John is tall”.

(ii) Single quotes:

  a. For mentioning words and sentences. e.g. ‘Green’ is a vague term. The sentence ‘John is tall’ contains the word ‘tall’.

  b. Nested Quotations: Austin claims that “‘Vague’ is itself vague”.

(iii) Italics:

  a. Used for what is expressed by an utterance: An utterance of the sentence ‘John is tall’ means *John is more than average height.*

  b. Initial introductions of technical terms. e.g. *Epistemicism* is a position in the philosophical literature.

  c. Mathematical and logical variables: x, y, F, G, ϕ, ψ, and functions: p, f.

  d. Titles of books/articles: In Williamson’s seminal work, *Vagueness*...

  e. Latin and other philosophical terms of art: *ceteris paribus, inter alia, tout court* etc.

(iv) Typewriter font: Logical and mathematical constants in formulas: $\texttt{green}(x)$, $\texttt{tall}(y)$.

(v) Small caps: Emphasis.

List of Figures

PART I

VAGUENESS AND TRUTH

1

Vagueness

1.1 Approaching Vagueness

Austin's observation that "'Vague' is itself vague" hints at how numerous the problems facing an account of vague terms in natural language are. Austin himself suggested a, presumably non-exhaustive, list of seven senses:

> "'Vague' itself is vague. Suppose I say that something, for instance somebody's description of a house, is vague; there is a quite large number of possible features—not necessarily defects, that depends on what is wanted—any or all of which the description might have and which might lead me to pronounce it vague. It might be (a) a rough description, conveying only a 'rough idea' of the thing to be described; or (b) ambiguous at certain points, so that the description would fit, might be taken to mean, either this or that; or (c) imprecise, not precisely specifying the features of the thing described; or (d) not very detailed; or (e) couched in general terms that would cover a lot of rather different cases; or (f) not very accurate; or perhaps also (g) not very full or complete... there is not just one way of being vague, or one way of not being vague, viz. being precise." (Austin 1962, pp. 125-6)

Of these seven, the term 'somebody's description' hides a further divide found in the literature on vagueness (one that mirrors a split in natural language semantics in general). Austin held that: "Usually it is USES of words, not words themselves, that are properly called 'vague'," but, at least with respect to his treatment of truth, Strawson and Searle accused Austin of conflating the act (the using of words) with what is expressed by the uses of words (something like the proposition expressed) (Strawson 1950), (Searle 1968). On that view, perhaps neither uses of words nor words are vague, but the propositions that (uses of) words express. Alternatively still, vagueness has been suggested to originate from objects rather than the language used to describe them.

More divisions in the literature abound:[1] Vagueness has been argued to be, amongst other things, a variety of ignorance (Williamson 1994), tolerance (Wright 1975), anything that gives rise to borderline cases, boundarylessness (Sainsbury 1997), and the having of blurred boundaries (Graff Fara 2000).

Common to a huge number of papers on vagueness and mentioned in almost every philosophical paper on the semantics of vague terms are *Sorites* paradoxes. Defining a Sorites is not straightforward, but they tend to come in three forms. Common to all three forms is a premise designed to be clearly acceptable/true and a conclusion designed to be unacceptable or false or *vice versa*. For example, '10,000 grains of sand make a heap' and '1 grain of sand makes a heap'. Between these two are one or more further premises. The *long Sorites* has, typically, a large number of indicative conditionals as premises where repeated applications of *modus ponens* takes one from the first premise to the conclusion. These conditionals are known in the literature as *tolerance conditionals*. The *short Sorites* comes in both a universally quantified and an inductive form. The former would, in the above case contain a premise such as 'For all n, if n grains of sand make a heap, then so do $n \pm 1$'. The inductive form of this would be, 'If n grains of sand make a heap, then so do $n \pm 1$'. These premises are known as *Sorites premises*.

Austin's methodological concern could be interpreted as being that, by discussing only some feature of terms affected by Sorites arguments we would not be discussing vagueness at all, but rather some philosophers' term, 'vagueness'. Not everyone views matters this way. Williamson defends allowing 'vague' to be an "artificially but legitimately restricted" technical term in philosophy (Williamson 1994, p. 71). He, somewhat unreasonably, states that describing the uses of 'vague' does not get at the vagueness of vague terms. The point is somewhat unreasonable because one way of beginning to determine in what way a term is vague and to what degree, whilst trying to avoid the danger of equivocation, is to pin down the different

[1] All references in this paragraph are mere examples. They are by no means exhaustive.

ways terms (or uses of them) can be vague in general. Doing so gives one a framework within which to classify uses of vague terms in order to be able to judge "in what respects" it is vague and "to what degree". For Williamson, 'vague' is only applied to cases of blurred boundaries where legitimacy arises from a need to face the challenge of Sorites paradoxes. Philosophy, like many other disciplines, might well be entitled to its technical terms, but Austin's methodological point should not be dismissed quite so lightly (even if the entirety of his methodology is not itself embraced in entirety).

Williamson was right descriptively. Vagueness in the philosophical literature certainly has become artificially restricted in various ways, both as a result of the assumptions writers have made and the conclusions that they have drawn. The literature on vagueness is also vast. This makes it hard to pin down where and how various writers are drawing theoretical lines at the outset of their theorising let alone by the end of it. Add to this mix the fact that, however vagueness is classified, it seems to be thoroughly ubiquitous, and you potentially get a fairly lethal concoction. The potential lethality can be expressed by way of a methodological suspicion. Treating 'vague' only as a technical term opens up the danger of delivering a warped result with regard to the conclusions we draw about what is vague, be it language, the world, or the sorts of things we can know. Put another way, the methodology of drawing fairly global conclusions about, say, the semantics of (possibly all) natural language predicates from fairly local phenomena like Sorites arguments, is suspect. Part of the goal of this work is to follow up on this suspicion.

The suspicion can be put another way. Sorites paradoxes have been used as the motivation for, and the success criteria of, some fairly hefty claims regarding, for example, what the meaning of vague terms (or uses of them) are like. This might be more than the Sorites paradox data are fit to bear. Better might be to find independently grounded reasons based on sufficiently broad-based data to draw conclusions about, say, the meanings of our words (and what is meant by particular uses of them). Having established these conclusions, Sorites paradoxes can be used as a test: is our picture of semantics one which leads us into paradox? A positive answer to this question would be a reason to think that we might have got something wrong on our bigger picture.

The main task of this chapter will be to highlight what kinds of assumptions about vague terms Sorites paradoxes have been taken to justify. First, in §1.2, this will require setting out the field of study in which I intend to work. This will not be the same as pinning down vagueness to a specific technical concept, rather it will determine what kind of questions to ask in the rest of the chapter. In §§1.3 - 1.7, we will look at some of the main treatments of vagueness that have been proposed. The main goal there will be to highlight the kinds of assumptions different theories make that pertain to the broad field of study defined in §1.2. Then, in §1.8, I shall draw together these different assumptions, provide a diagnosis of why certain disputes in the literature seem to have become fairly irresolvable, and present a methodology for arriving at a treatment of vagueness that is more in line with the Austinian suspicion given above.

1.2 Defining the Field of Study

There are, broadly, three types of phenomena that vagueness has been held to be: a metaphysical phenomenon, a semantic phenomenon and an epistemic phenomenon.[2] The difficulty that faces us is that the number of assumptions and conclusions drawn by all accounts of vagueness would be too broad and wide ranging to draw our own conclusions from. We must delimit the kinds of assumptions and conclusions that will be considered without biasing some treatment of vagueness over another. It would not do to take only certain domains of assumptions into consideration, and, in doing so, exclude certain approaches to vagueness from being considered on the basis of this choice.

The task is made simpler since this will be an investigation into the vagueness of natural language terms (or uses of them). Even if one thinks that the original source of vagueness is something non-linguistic (metaphysical or epistemic), reaching this conclusion will require taking a semantic stance. One could conduct a study of vagueness that took as its field of study, how we can know such and such or what the underlying nature of things are. By focusing on language via the uses and meanings of terms, we will not exclude the viability of epistemic or metaphysical accounts, or so I shall argue.

[2]Recently, some (for example Schiffer (2003)) have defended the view that vagueness is psychological. I will not consider such a view here although some aspects of psychology with respect to vagueness will be considered in §1.5.1.

1.2.1 Metaphysics

The metaphysical aspects of vagueness, if there are any, will be shelved in our subsequent discussion. First, we are interested in what makes a natural language term, or perhaps, a particular use of one, vague. It may well be that whatever answer we provide to that question will turn out to supervene on some heavy duty metaphysics regarding the individuation of objects or properties, or on some facts about whether objects themselves have sharp individuation, but those kinds of questions should not prevent us from talking about language and language use. For example, say that we come to the conclusion that some use of a vague term is indeterminate. This view is compatible with two incompatible metaphysical viewpoints. It could be that the properties in the world that the term speaks of, or was used to speak of (if that is what terms do speak of) are fully determinate. From that position, perhaps indeterminacy would be held to arise due to something semantic (say, that the reference of the term was indeterminate in some way). Alternatively, it could be that the property is fuzzy and indeterminate in some way, even if the reference of the term used to refer to this property is completely fixed. The point is best demonstrated with an example. It should be made clear that in the following, no assumptions are being made about what terms do refer to, only that metaphysical positions on objects do not need to be considered in order to speak about language and what terms can (be used to) mean.

A classic example used to discuss the metaphysical view is the problem of the many (Geach 1962; Unger 1980) which was evocatively described by Lewis as the paradox of the 1001 cats (Lewis 1993). A cat is an object and this object has many parts (paws, whiskers and claws etc.). Among the cat's parts are all the hairs that make up its fur, but the (meta)physical identity of the cat seems to show tolerance to the number of hairs it has. A cat with no hairs removed is just as good a candidate for being that cat as one with a single hair or multiple hairs removed. Indeed, were the cat to suffer from a form of feline alopecia, there might still be a good case for thinking that Felix would still be Felix. The same sort of line can also be taken towards other parts of the cat, or towards hairs that are half or one third or three quarters fallen out. So we have both one cat, but also, perhaps, many slightly meta(physically) different cats.[3] Whereas the semantic view might hold that 'Felix' is vague, the metaphysical view holds that Felix is

[3] By which it is meant that there are many candidates for which physical object IS Felix, each of which could itself count as a cat.

vague. However, there are two ways the metaphysical view could be filled out. Either there is one cat: ‘Felix’ refers to Felix, but the object itself is vague, or it is vague which of many sharp Felix contenders ‘Felix’ refers to. On the first view there is no fact of the matter how many hairs Felix has, as Felix’s (meta)physical form is fuzzy when it comes to numbers of hairs. On the second, every Felix-contender has a precise number of hairs, but it is indeterminate which contender is Felix.

In principle, the two types of vagueness (semantic and metaphysical) could both be in effect such that the term ‘Felix’ would have borderline cases and the object Felix would have no determinate extension (either because the object is vague, or because there are multiple Felix-contenders). Then, there would be situations where we could identify, at least conceptually, indeterminacy (or uncertainty) arising from the vagueness of the term, from the vagueness of the object, and from both.

Since both the the vagueness of objects and the vagueness of terms can be sources of indeterminacy (or uncertainty), it is acceptable to study the two independently. This is because the two are conceptually independent: Given that vagueness is a phenomenon that requires an explanation, (i) An argument for the vagueness of objects is not an argument against the vagueness of terms, nor is an argument for the vagueness of terms, an argument against the vagueness of objects. (ii) An argument against the vagueness of objects is not an argument against the vagueness of terms, nor is an argument against the vagueness of terms an argument against the vagueness of objects.[4] Therefore, even if metaphysical questions of vagueness have an answer (knowable or otherwise), we are able to pursue a study of vagueness in language without answering them. Next, I will suggest that the view that vagueness is held to arise as a result of the meanings of terms is by no means inconsistent with holding that vagueness arises as a result of what we do or don’t know.

1.2.2 Epistemology

Epistemic positions should be kept distinct from *epistemicist* positions on vagueness. *Epistemicism* is the doctrine that vague terms have, semantically, sharp boundaries that are unknowable (full details of this view will be given in §1.4). For every term, or perhaps every concept which a term expresses, there will be, relative to some context, a determinate answer

[4]Although, if terms and objects exhaust the options for what can be vague, an argument against one could be an argument for the other. I will not pursue this strategy.

to whether something (anything) falls under that concept and/or is in the extension of the term. Vagueness, in the sense of borderline cases, arises because we cannot know where these boundaries are situated. Not all epistemic positions need to be epistemicist. One might view vagueness as relating in some way to knowledge (or belief) which can be partially defined by ignorance, an absence of knowledge, or certainty about something. However, that thing that we are uncertain or ignorant about need not be the extension of the term (where to locate its boundary).

Whether epistemicist, or merely epistemic, semantic questions must be answered before epistemological ones.[5] The reason for this will be made clear below. In brief, I will argue that the kinds of things that we are ignorant of/uncertain about will be determined by the semantic assumptions we make. Hence, in which respect an account of vagueness is epistemic will be determined by what we first say about semantics.

Suppose we think that part of an account of vagueness is about ignorance of X (or alternatively uncertainty whether X). At least part of our reasons for thinking this, given that we are ignoring metaphysical questions, will have to relate to language (whether it is words, uses of words, what words express, or even what our grasp of what words express). It may well be that we think that we don't know what exactly has been expressed/said. In that case, we will, at the very least, be ignorant over that. It may be that we know/grasp exactly what (uses of) words mean or express. In which case, what we are ignorant of will be determined by what was meant/expressed. This is because what was meant/expressed will leave us not knowing certain things (those words would not tell us THAT).

For example, epistemicist positions are based on fairly conservative claims over the extensions of terms. These conservative views over extensions are exactly what feed into the claim that we are ignorant about where the extensions, posited by the theory, are. Other epistemic (although not openly epistemicist) views employ uncertainty rather than ignorance ((Lassiter 2011), (Frazee and Beaver 2010)). On these views, linguistic terms have certain semantic features. For Frazee and Beaver (2010) this is a contextually determined threshold that sets a standard within a comparison class. For Lassiter (2011), uncertainty is generated from the possibility of various (sharp) languages being the ones being currently used. In either case, what it is we are uncertain of falls out of the underlying semantics for the terms themselves.

[5]It is nonetheless possible that some epistemic questions may themselves be taken to be semantic questions.

Hence the field of study we can focus on will be, broadly speaking, semantic. Theories will be classified by the commitments they make on semantic grounds. This will include, not only meanings, denotations, relations to truth-values and the like, but also, where necessary, some more pragmatic factors. For example, what terms are used for, and how, if at all, meanings, denotations etc. change with context.

The next five sections will survey different approaches taken in the literature. Some of the general promises and problems of these approaches will be discussed, but primarily, I will try to draw out the semantic commitments of each position. Some of these commitments will be assessed in this chapter. A general critique of truth conditional semantics will be made in chapter 2.

1.3 Vagueness in the Literature – Overview

The vagueness literature is pretty vast, and so any overview is going to have to find a way to divide the literature into different kinds of approaches to particular problems. There is no single right way to make these divisions. Since our interest is in the semantic commitments that each account of vagueness makes, I will divide the literature up along semantic lines. In this section, I will give a broad, non-theory-specific overview of the different approaches to vagueness to be found in the literature.

The position one adopts in the vagueness literature could be seen as the result of making a range of choices about semantic assumptions. Figure 1, below, gives a rough picture of how such choices can lead to a particular view. A glance at the tree reveals a clustering of theories based on common mothers. The first binary decision reveals two categories of theories. The daughters of node [1] are based on a classical conception of a proposition (as something like a function from indices to Boolean truth values). The daughters of [0] adopt a non-classical account of propositions. The binary decision at node [1] marks a division within classical proposition accounts. The daughters of [11] explain the meanings of sentences/utterances containing vague expressions using a single classical proposition. Whereas, the daughters of [10] appeal to multiple propositions. These divisions carve out three groups of accounts of vagueness that I will give examples of. In order of discussion they are: (i) terminal daughters of [11]. (ii) terminal daughters of [10]. (iii) terminal daughters of [0].

Our first category (daughters of [11], bottom left cluster) contains theor-

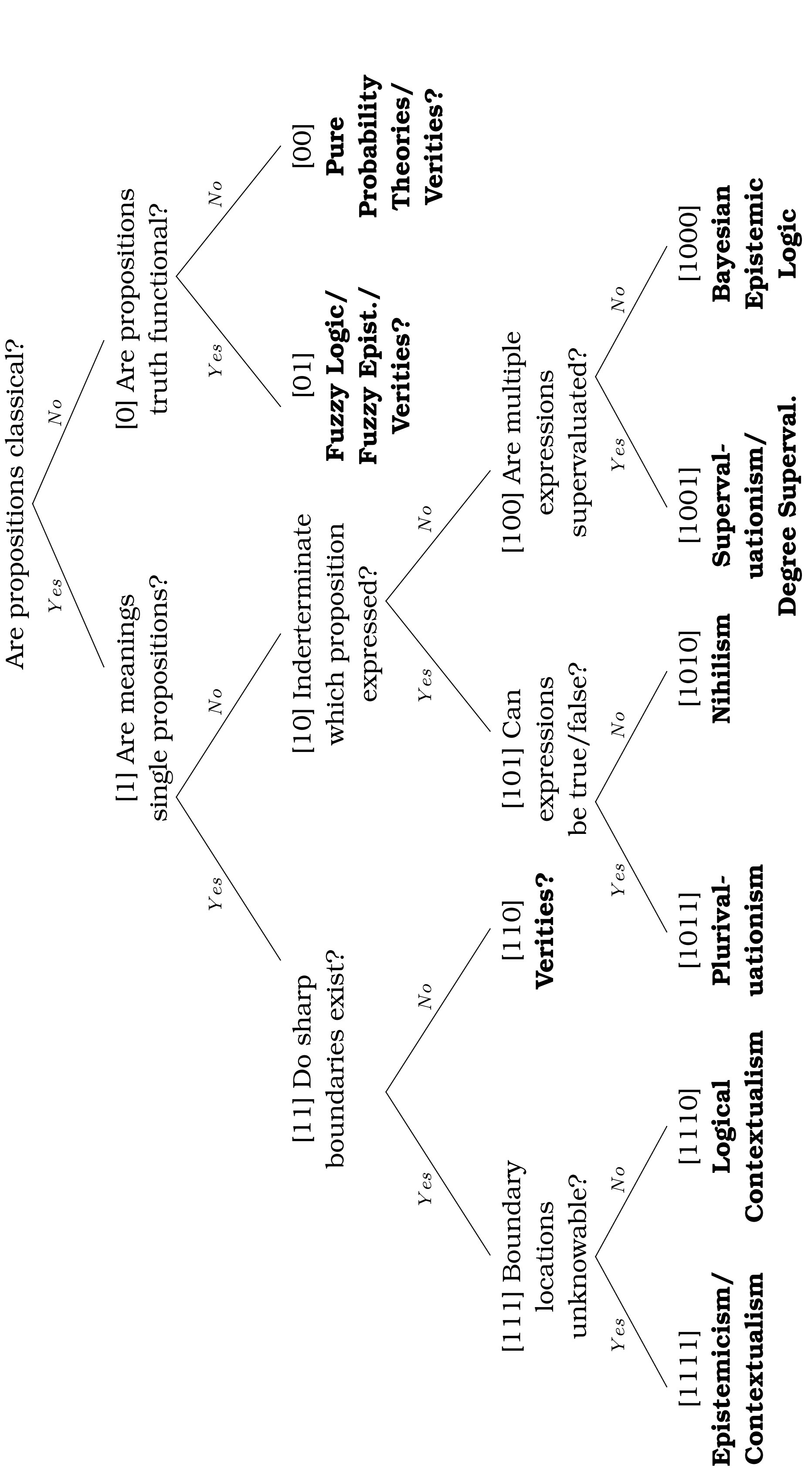

Figure 1 - Semantic Commitments of Theories of Vagueness

ies that associate meanings of utterances with SINGLE PROPOSITIONS, where utterances are either true or false *simpliciter*. The most semantically simple amongst these is epistemicism. Also falling into this category are contextualist views. Contextualist positions can be thoroughly classical with regard to content and truth-value, but relativise truth-values and/or extensions to contexts. Epistemicist positions are free to do this also. The difference between them is that epistemicism addresses Sorites cases via ignorance of boundaries, whereas contextualism addresses them via minor changes in context mid-argument. The outlier in this group sits slightly higher in the tree and its position there is marked with a query. Indeed, the same theory is also marked on the other side of the tree. Dorothy Edgington's *verities* account can be read fairly classically. On this reading, there are still only two truth values, but degrees of closeness to clear truth (verities) help us, for some purposes, to address gradience in our semantics in a way that speaking only in terms of absolute values cannot. There is, however, another reading (not necessarily Edgington's) on which verities are taken to be something more akin to degrees of truth. Some discussion of verities will be made in §1.7, but a full discussion will not be conducted until chapter 8. There, I will argue for the reading of verities as relating to uncertainty (node [00]).

The second broad category of theories (daughters of [10], bottom middle cluster) are slightly more complex. These theories associate meanings of sentences/utterances with MULTIPLE PROPOSITIONS (they differ on the status of these propositions and over whether they are all expressed). This broad category includes theories such as supervaluationism, plurivaluationism and nihilism.[6] The difference between the previous category of theories and this one is that the first adopt a traditional stance towards truth as well as content. A proposition is expressed and it is either true or false. Multiple proposition theories are forced to say more about truth. If an utterance really expresses more than one proposition, each of which is true or false, then truth of the utterance may be defined over the truth-values of the propositions expressed (supervaluationism), or one may deny that any utterances containing vague terms are ever literally true or false (nihilism).

Theories in the third category (daughters of [0], top right cluster) abandon (classical) propositions. These DEGREE THEORIES model meaning via some kind of more graded entity. Good cases in point are fuzzy views. Rather than expressing a multitude of classical propositions, fuzzy approaches suggest

[6]At least of the sort of nihilism defended by Braun and Sider (2007).

that meanings can be represented by fuzzy sets and that propositions have fuzzy truth conditions. Such views may or may not posit that fuzziness in the language is matched by fuzziness in the world. On fuzzy views, utterances receive non-classical truth-values which are usually conceptualised as degrees of truth. However, not all degree theories are fuzzy. An alternative to fuzzy degrees of truth is to characterise meaning directly in terms of probabilities. Probabilities represent something like degrees of belief. It is a version of this kind of theory that I will defend in this thesis.

As can be seen from the overview above, one can be classical with regard to content and truth-values, classical with regard to content but not truth value, or classical with regard to neither.[7] Potentially, one might also be classical with regard to truth values, but not with regard to content. However, the coherence of such a view would turn on what the non-classical content was like.

1.4 Vagueness in the Literature – Epistemicism

Epistemicism, according to one of its main proponents Timothy Williamson, has it's roots in classical Greek Stoicism. As a philosophical treatment of vagueness it saw a re-emergence that started in the late eighties with the work of Roy Sorensen (Sorensen 1988), (Sorensen 2001), but it became a tour de force in the 90s with Williamson's seminal book, *Vagueness* (Williamson 1994).

Epistemicism should not be characterised as the position that vagueness is an epistemic (as opposed to a semantic) phenomenon. As suggested in §1.2.2, it is possible to defend an epistemic approach to vagueness that is not epistemicist. It is better to view epistemicism via the claim that the unacceptability of sharp boundaries for vague terms can be explained as the (necessary) ignorance of those boundaries. Epistemicists present a dilemma of two seemingly unacceptable options: Accept that vague predicates have sharp boundaries, or cease using classical logic in formal semantics to model meaning and reasoning in natural languages. Epistemicism opts for the former.

In their favour, Williamson and Sorensen seem to share the suspicion detailed above that Sorites arguments should not be (mis)used to motivate far-reaching conclusions about semantics. As a result, they defend a

[7]Notice that this is not a claim about logic. As we shall see, being non-classical with relation to meaning and truth does not mean that one must deny any theorem of classical logic.

traditional stance towards truth and meaning. This leads them to find explanations of vagueness in an epistemic domain. In the next few chapters, I will consider the merits of this traditional stance on independent grounds.

1.4.1 Sorensen's Epistemicism

Sorensen (1988) argues that it is a mistake to "shrink the domain of truth down to the domain of the knowable". In other words, we ought to get used to the fact that there are truths that we do not and cannot know.

Sorensen draws a parallel between the lottery and Sorites paradoxes (Sorensen 1988, pp. 239-41). The lottery paradox is a paradox about knowledge and/or belief. In my opinion, it is most forcefully made in relation to reasonable belief. One can reasonably believe that one particular ticket will not win the lottery, and one can reasonably believe the same of two particular tickets. Yet clearly, one cannot reasonably believe this of all tickets. The question is, where does reasonable belief stop? In the lottery paradox, we have some kind of internal inconsistency in that we rationally believe that some ticket will win and rationally believe of no particular ticket that it will win. The effect is that we hold a collection of inconsistent, although rational beliefs, one of which must be false. In a long Sorites, instead of having a universal or inductive Sorites premise, we instead have multiple tolerance conditional premises:

1. 10,000 grains make a heap
2. If 10,000 grains make a heap, 9,999 grains make a heap
3. If 9,999 grains make a heap, 9,998 grains make a heap

.

.

10,001. 1 grain makes a heap

According to Sorensen, we can rationally believe all of the premises of the Sorites even though they are incompatible, given that their conclusion is false.

Sorensen (2001) lays out what he terms 'the basic argument'. Take a Sorites, like the one above, but where all of the intermediate premises are replaced by an induction step (if a collection of n grains of sand is a heap, then so is a collection of $n-1$ grains). The induction step seems to be the only candidate one can reasonably reject. But the negation of the induction step implies that one grain of sand can make a difference. So there must be a sharp threshold for 'heap'. Having provided the basic argument, Sorensen spends the rest of (Sorensen 2001) attempting to minimise the discomfort that such a conclusion generates.

The basic argument is given in natural language, and in natural language it has a pull. Sorensen states that what guides his attraction to the basic argument is allegiance to "lightweight, common sense, textbook logic" (Sorensen 2001, p. 1). The denial of, 'if a collection of n grains of sand is a heap, then so is a collection of $n-1$ grains', is an assertion of its negation.[8] According to lightweight, common sense, textbook logic, its negation is true only if one grain makes a difference.

In order to get the basic argument to work, elementary logic of the kind we learn as first-year undergraduates isn't enough. That kind of logic tells us that if we start with truths, then we get truths after applying specified inferences to formally defined sentences in the logical language. Classical Boolean logic is completely undefined for anything other than Boolean truth values. This means that non-classical conceptions of truth and logic are being assumed not to be applicable. This rules out many alternatives, seemingly by fiat. For example, it rules out intuitionistic logic (which blocks the simple argument).[9]

Perhaps Sorensen does not mean to apply 'logic' in a formal sense, but intends to make use of some pre-theoretic common sense notion of inference. In which case, it must then be asked of him to specify what a grasp of this logic entails. If it entails that, say, all sentences have one of the two Boolean truth values, and that all things are either F or not-F, then the above point still holds.

Applying everyday textbook logic to natural language requires some translation. One can have an allegiance to textbook logic, and yet still be entitled to suggest that natural language sentences diverge in various ways from their classical logic analogues. The literature on indicative conditionals is testament to how unsatisfying simple material implication interpretations of 'if..., then,...' are. Sorensen assumes that simple logic is directly applicable to all (relevant) elements in the simple argument. In doing so, he makes some conservative assumptions about semantics. He assumes that simple classical propositions can be assigned as interpretations of the premises and conclusion of the simple argument. He further assumes that

[8]This assumption is disputed in the literature. For example, (Tappenden 1993) and (Richard 2000) argue in favour of metalinguistic negation (under which one asserts the incorrectness of asserting ϕ without asserting $\neg\phi$).

[9]See (Putnam 1983) for a proposal applying intuitionistic logic to the Sorites. See (Schwartz and Throop 1991) for a response.

these interpretations ensure classical Boolean values for all such sentences. Both of these assumptions amount to a ruling out of most alternatives. The simple argument is not so much an argument for epistemicism as a statement of it.

As well as Sorensen's position on logic and meaning, another interesting part of his view concerns truth. Sorensen's epistemicism is committed to the view that all declarative utterances are either true *simpliciter* or false *simpliciter*, but more subtlety enters the picture when we allow that being true or false *simpliciter* can be understood in different ways.

One way to view epistemicist truth is as a mere device of disquotation. On disquotational views, the addition of 'is true' to a sentence merely provides semantic descent back to the assertion that would be made by the use of that sentence. Indeed, Sorensen suggests that the semantic thinness of a deflationary conception of truth leads fairly inexorably into epistemicism.[10]

Sorensen also argues that epistemicism is compatible with a more inflationary approach such as truth-maker theory (Armstrong 1989).[11] However, his version of truth-maker theory is a pretty far cry from traditional formulations of it. Sorensen defends the notion of an *underspecific truth-maker*. This leads him to support the notion of a *truth-maker gap*. Truth-maker gaps imply that some statements can be autonomously true (Sorensen 2001, p. 177). A serious challenge Sorensen must face is whether he can keep his dichotomous view of truth from collapsing. If some statements are truth-makerlessly true, why can't all be true (whether they have a truth-maker or not)? Furthermore, even if a statement lacks a truthmaker, this may still require a semantic story to be told. How can we determine whether or not a statement is a gappy truth (falsity) or one with a truth (false) maker? We still need to know what statements mean in order to determine whether some state of affairs lacks a truthmaker for our statement or not.

In the next chapter, views on truth will be considered in some detail. In particular, we will assess the viability of appealing to truth (and truth conditions) in order to provide a semantic theory.

[10]Sorensen cites Horwich (1990) as an example of this. The section on vagueness in Horwich is in effect an early version of the argument Williamson went on to make about the retention (or, at least, the non-denial) of the principle of bivalence. Williamson's argument will be discussed below.

[11]The main principle for truth-maker theory is that all contingent truth must have something in the world that makes them true (Armstrong 1989, p. 88).

1.4.2 Williamson's Epistemicism

Two lines of thought drive Williamson's account of vagueness. One is based on the principle of bivalence (that every utterance is either true or false):

> "Suppose that an utterance of 'TW is thin' is either true or false. Then since we do not know *that TW is thin* and do not know *that TW is not thin*, we are ignorant of something." (Williamson 1994, p. 185), (my italics).

The other is based on pragmatic considerations relating to the abandonment of bivalence:

> "If one abandons the principle of bivalence for vague utterances, one pays a high price. One can no longer apply truth conditional semantics to them and probably not even classical logic. Yet classical semantics and logic are vastly superior to the alternatives in simplicity, power, past success, and integration in other domains. It would not be wholly unreasonable to insist on these grounds alone that bivalence MUST somehow apply to vague utterances, attributing any contrary appearances to our lack of insight. Not every anomaly falsifies a theory. That attitude might eventually cease to be tenable, if some non-classical treatment of vagueness were genuinely illuminating. No such treatment has been found." (Williamson 1994, p. 186)

In the first quote, Williamson tacitly relies upon disquotational principles to move from "'TW is thin' is true or 'TW is thin' is false" to identify *that TW is thin* and that *that TW is not thin*, respectively, as things to know or be ignorant of pertaining to the truth or falsity of 'TW is thin'. This connection is then used by Williamson to do some philosophical work. Part of that work is in his argument from a denial of a specific instance of bivalence to a contradiction.

The Argument from the Denial of (an instance of) Bivalence

At the heart of (Williamson 1994) is Williamson's argument against the cogency of a denial of a specific instance of bivalence.[12] The argument

[12] The 'specific instance' condition means that the argument does not apply to intuitionism.

begins with a discussion of what the relevant notion of bivalence is when discussing language. Bivalence could be held to apply to all and every proposition expressed by our utterances.[13] This kind of bivalence is not where Williamson thinks we ought to start:

> "In a relevant form of bivalence, the truth bearers are (perhaps with a little artificiality), the utterances themselves. The principle is explicitly restricted to occasions when someone uses an utterance to say that something is the case, in brief (if again with a little artificiality), when the utterance says that something is the case." (Williamson 1994, p. 187)

Here Williamson seems to be influenced by use theories of meaning (Austin 1950/1979) (Wittgenstein 1953/2009). He is assuming that utterances express unique propositions (and are thereby true or false). What proposition an utterance expresses will presumably not turn only on the sentence that is being used, since Williamson accepts that what is expressed (by the use of a sentence) can change with context. For Williamson, utterances are truth bearers and, if they express anything at all, utterances express single propositions.[14] Single propositions are classical. Therefore, what is expressed by an utterance can be represented as some classically truth-evaluable expression in the language of representation/the metalanguage.[15]

With this view of utterances assumed, Williamson gives a principle of bivalence for utterances (B):

(B) If u says that P, then either u is true or u is false.

u is to be replaced by the name of an utterance in the metalanguage and P should be replaced by a sentence in the metalanguage that expresses what u says. With Tarski in mind, Williamson then provides two schemas:

(T) If u says that P, then u is true if and only if P.

(F) If u says that P, then u is false if and only if not-P.

[13] If propositions are defined as classical functions into Boolean truth values, then bivalence will apply to propositions.

[14] Cases of semantic/syntactic ambiguity are not excluded. Williamson allows that more than one speech act could be performed simultaneously. This would count as two (or more) utterances each expressing a single proposition.

[15] Or, at least, that representing things that way is the simplest, most powerful, most successful way of doing so (see above).

Suppose that utterance u says that P, but is neither true nor false. u would be a counterexample to (B). By *modus ponens*, 'u says that P', (T), and (F) give:

(T′) u is true if and only if P.

(F′) u is false if and only if not-P.

A denial of (B) involves asserting the following:

(Not-C) Not: either u is true or u is false.

(T′) and (F′) are statements of logical equivalence and so 'u is true' and 'u is false' can be replaced with 'P' and 'not-P' in (Not-C). Application of one of De Morgan's laws yields a contradiction:

(C) Not-P and not-not-P.

Williamson concludes that denying a specific instance of bivalence yields a contradiction. He takes any account that yields a flat contradiction to be a *reductio* on that account. Any modelling of vague predicates that entails the denial of an instance of bivalence then ends up in contradiction.

A few comments about the target of Williamson's argument should be made here. He does not claim to be arguing against intuitionism. He takes intuitionists to be denying the general principle of bivalence whilst neither denying nor asserting a specific instance of the principle (where one does not have a proof of either disjunct) (Williamson 1994, pp. 192-3). However, it is insufficient for Williamson to simply note that intuitionism is not targeted by his argument. Intuitionism is, itself, a counterexample to some key claims that Williamson uses to motivate epistemicism. For example, classical logic and semantics are not vastly superior to their intuitionistic counterparts.[16] Intuitionists may still have to say something about vagueness.[17] However, if intuitionism is a way of avoiding epistemicism, Williamson should say more about it. In the account I will develop, I will make use of classical probability theory and classical logic. I take it that nothing I say would inhibit a move to intuitionistic alternatives.

With regard to (T) and (F) we are told:

[16] See the quote at the beginning of §1.4.2.

[17] It is not clear that intuitionism alone solves the Sorites. See (Putnam 1985) for some references and a brief summary.

> "There may not be much more to be said about truth and falsity than (T) and (F) say, but they do at least seem central to our ordinary understanding of those notions". (Williamson 1994, p. 188)

But appeals to an ordinary understanding of those notions are not useful here since the very point is to draw a conclusion regarding the relations between truth, falsity and utterances on some, possibly non-ordinary, understanding of truth. But this appeal hides a more serious point: (T) and (F) themselves admit of different interpretations. Here are some examples:

(I) The interpretation of the status of the T-schema: P is supposed to be a sentence of the metalanguage which has a truth-value (under some conception of truth). This is Tarski's conception. It is widely accepted that Tarski's T-theory, as read, is not directly and simply applicable to natural languages. Montague's claim that English is a formal language was more an expression of optimism that, with further research, a fragment of English could slowly be expanded to give a representative semantics for a whole language. Understood as Tarski meant it, T-theories applied to natural languages carry a substantive claim regarding the viability of applying formal semantics to such languages.[18] Not everyone will want to make this claim. Just how the semantics of the object language relate to derivations of instances of (T) and (F) may, therefore, carry substantial assumptions about truth itself.

(II) The interpretation of 'true': Irrespective of whether one supposes that instances of the T-schema must be formally derivable from a semantics for that language, we can ask in what sense we are to understand 'true'. Disquotation and the Law of Excluded Middle (LEM) can be used to derive bivalence. So the way that 'true' and 'false' can or cannot be used in disquotation may determine whether or not one accepts bivalence, or, at least, how one should understand a statement of bivalence. A more rigorous discussion of this approach to bivalence and disquotation will be given in §8.3. There, I will suggest that a degree theorist can accept versions of disquotation and versions of bivalence. This is not a problem with Williamson's argument, but it suggests that one can accept bivalence, without accepting Williamson's interpretation of it.

(III) The interpretation of 'if and only if': Richard (2000) has suggested that Williamson's argument assumes that T-schemas must be set out extensionally (that the 'if and only if' must be read as a material biconditional).

[18]Tarski himself was famously sceptical of the possibility of this project.

Richard sets out an intensional version of (T) and (F) upon which Williamson's conclusion does not follow. In brief, the proposal is that (T) and (F) should be read as saying that their left and right hand sides are true in all and only the same worlds. If, as Richard claims, the more plausible readings for (T) and (F) are intensional readings, then is is plausible for an utterance to be meaningful while still lacking a truth-value at some indices. This means that the intensional versions of (T) and (F) can be true whilst their extensions at some index, and the corresponding instance of bivalence, can lack a truth value. The argument is therefore blocked.

Certain other assumptions can be drawn out of the above. Williamson's writing suggests a certain sympathy for the viability of an extensional, Tarski-cum-Davidson inspired approach to truth, possibly with an extensional and decidedly classical sounding semantics in the background (with which instances of the T-schema can be derived). Further, such theorising about truth seems limited to an (ordinary) understanding of truth, which precludes there being some understandings of truth on which (T) and (F) could be held to fail. From these assumptions, an interrelated view of truth and meaning is formed from which epistemicist conclusions start to flow. It was already established that he supposes utterances to express single propositions. It now appears that the propositions they express are being assumed to be capable of playing a role in an extensional classical semantics (that supports Tarski's conception of (T) and (F)). This allows one, by way of instances of (T) and (F) schemata for utterances, to get from an assertion of bivalence to a commitment to sharp boundaries.

The Role of Context in Epistemicism

There is one final part of Williamson that I will consider. This concerns his positive view of why it is that we should be ignorant of sharp boundaries for vague terms: Inexactness leads to knowledge only within certain margins for error. If I believe that the room I am in is exactly 10m long, but my ability to judge room lengths is reliable only to the nearest half meter, then, even if I am right about the length of the room, I do not know that the room is exactly 10m long. This is because, under Williamson's epistemology, knowledge must support certain counterfactuals.[19] In brief, the room could have been not-exactly 10m long, say 10.05m long or even 10.49m long. If

[19]For a detailed elaboration on the counterfactual principles underlying Williamson's epistemology, see (Kearns and Magidor 2008).

in such situations I would still have believed the room to be exactly 10m long, then I cannot know it. My belief is too risky.

Williamson applies this notion to vagueness. For regular knowledge, inexactness has its source in things like perceptual inexactitude. In the above case, I didn't know that the room was exactly 10m long, but I could, by the use of a laser measuring device, have come to know the length of the room within some much narrower margin for error. What interests us here is not Williamson's epistemology. We want to know his assumptions about language. In his account of vagueness, Williamson locates inexactitude in our ability to determine meanings:

> "What distinguishes vagueness as a source of inexactness is that the margin for error principles to which it gives rise advert to small differences in meaning, not to small differences in the objects under discussion." (Williamson 1994, p. 231)

Unlike Sorensen, Williamson defends a use theory of meaning. The meanings of vague terms supervene on how they are used. Nothing but use can stabilise the meanings of vague terms. Slight variations in the way a term is used inexorably lead to a slight difference in its meaning.

It is not entirely clear how Williamson wants this sensitivity to use to operate. Earlier in *Vagueness* he admits of possible context effects on vague terms. Vague terms have sharp boundaries, but not necessarily the same sharp boundaries in every context. Williamson seems to associate meanings of terms, in some way, with their extensions:

> "A slight shift along one axis of measurement in all our dispositions to use 'thin' would slightly shift the meaning and extension of 'thin'. On the epistemic view, the boundary of 'thin' is sharp but unstable." (Williamson 1994, p. 231)

He does not tell us if there is a relation of supervenience between meaning and extension, but the above passage at least tells us that extension must also, in some way, supervene on use. We may, then, wonder whether two processes operate on extensions of terms. In one sense, contexts of use may change the extension of a term. In another sense, all our histories of use of that term, might, independently of contexts of use, also determine a term's extension. Then again, at times, Williamson seems to speak as though it is a history of use that determines what extension a term will have in a current context. Perhaps it is enough to characterise Williamson's position in the following way: Relative to some context, broadly understood as a combination of past linguistic usage and current facts about the situation of use, a vague term has a specific extension.

In this chapter, I leave aside whether this may, or may not, seem like a plausible view to adopt. In the next chapter, I suggest it is not. For now, I wish only to mark that some positions on vagueness take as completely central the view that, under a broad notion of context, terms have a sharp extension that can vary with context.

1.5 Vagueness in the Literature – Contextualism

A variety of contextualist positions are to be found in the literature (Kamp 1981), (Soames 1999), (Raffman 1996), (Graff Fara 2000). Although some of these are fairly simple, semantically speaking, others are more complex. Here, I will focus on the simpler ones. In particular, I will not address some of the intricacies of Kamp's account.[20] Two accounts will be considered. Raffman's internal context approach and Graff Fara's interest relative approach.

The broad idea behind contextualist accounts is that what is meant by a vague predicate, or perhaps what property they refer to, can change and vary in a dynamic way. The dynamic part is important. An easy but unfair objection to a contextualist account of vagueness would be that vagueness (be it susceptibility to Sorites arguments, blurred boundaries, borderline cases, or whatever), arises even when the context is fixed: Even relative to a single context, vagueness remains. For example, relativising 'tall' to a context of basketball players may raise the standard to count as 'tall', but that relativisation would still yield borderline cases. Perhaps I, at a ordinarily lofty 6ft 2in would be a borderline case of tall for a basketball player. The objection is unfair in light of a dynamic approach to context and in light of the aims that contextualists see themselves as pursuing. A treatment of the Sorites paradox lies at the centre of most contextualist accounts of vagueness. They seek to account for the inconsistency of soritical reasoning in terms of shifts in context that occur as the reasoning progresses.

[20] Kamp's contextualist solution is in some sense fairly traditional. However, he makes adjustments to the classical interpretation of universal quantification and the material conditional which allow him to move from a coherent context to an incoherent one by the incorporation of sentences which are true by the lights of the former. Although interesting, a full analysis of the reaches of such changes will not be made here.

1.5.1 Raffman's Internalism

Diana Raffman ((Raffman 1994), (Raffman 1996)) has proposed that the key to providing an account of vagueness (by which she tends to mean solving the Sorites paradox) is to pay attention to the psychological changes that, she claims, we can infer are going on in agents as they go down a path of soritical reasoning. Raffman detects the presence of a psychological factor via certain principles relating to linguistic competence. One premise is that, when being marched through a Sorites argument, every competent speaker must reach a point where they refuse to apply the predicate in question. Furthermore, since individuals differ between where they dig in their heels (from each other, and from themselves over time), and since it is senseless, even relative to a context, to call any one speaker mistaken because they have stopped at a slightly different place than another, Raffman concludes that the correct application of a vague term varies from person to person and time to time even relative to a context[21].

This view of the correct application of a predicate, plus data regarding personal and interpersonal differences in application behaviour, lead Raffman to the hypothesis that every individual has an internal context relative to which we are disposed to make particular judgements about whether to apply a predicate or not. When a speaker digs in their heels on a Sorites slope, Raffman contends that a gestalt shift has taken place in their internal context (effectively, their internal sub-personal state). For every one of these states there is associated a predicate, P, and a set of three dispositions: The disposition to judge some set of things P, the disposition to judge some exclusive set of things as not-P, and, for some one thing that lies in between the things we are disposed to classify as P and the things we are disposed to classify as not-P, we have a disposition to switch into another state when we consider that thing. Raffman treats this in-betweener as being classified as a non-P item relative to the internal context. As such, bivalence needn't be abandoned on her view.

The opacity of the nature of these internal states is supposed to be an advantage of this kind of view. We are all trapped in some sort of internal state and we can never be sure what will switch us into another one (other than by testing it against a Sorites series). As soon as we get too close to the boundary and start to consider the in-betweener case, a switch happens and the boundary of the new state is, once more, far away.

[21]Here 'context' is being understood as sensitivity to a comparison class. For example, I am tall for a philosophy student, but not tall for a basketball player.

In some sense Raffman's internal contexts can be seen as the tokening of concepts that are extensionally defined.[22] The concept we are employing is (modulo external context) in a one-to-one relationship with the state we are in. As soon as we try to employ the concept near to its boundary we unconsciously switch to another concept with different boundaries.

One may or may not feel comfortable with this kind of psychological story. One immediate concern is that it is hard to see how it could be falsified, even in principle. Such concerns are not ours. Important to note is that behind the psychologism is a completely classical account. Relative to some internal/external context pair associated with some predicate P, everything will be P or not-P even if we are never in a position to judge it so.

1.5.2 Graff Fara's Interest Relativity

Delia Graff Fara's account of vagueness (Graff Fara 2000) is the first we shall consider that makes some substantive and non-traditional claims about the meaning of vague terms.[23] Like Raffman's, Graff Fara's account of context-sensitivity incorporates psychological elements to account for vagueness beyond the external, tall-for-a-basketball-player, type of considerations. Graff Fara's account is, however, framed in more familiar terms than Raffman's hidden states. The basic idea is that aims, intentions and desires play a role in whether or not something counts as P – the question as to whether something is P can depend on what our interests are in the matter. For example, whether or not something is heavy, even heavy for a paperweight, may depend on what I am trying to do. If I desire to keep my papers from blowing away in a very windy environment, something that is heavy for a paperweight might not be heavy enough for my purposes. In such an interest-relative context, whether something counts as heavy will depend on whether it is heavy enough to do the job.

The two types of context-sensitivity, call them 'norm relative' and 'interest relative', are combined together to give an analysis of vague expressions in terms of *typicality* (which is norm relative), and *significance* (which is interest relative). The meaning of 'is tall' is not just a function from individuals to truth values. 'Tall' stands for the relational property of *being significantly greater in height than is typical.* Whether or not I am greater in

[22]Raffman is ambivalent about whether the extensions of internal states should vary or be fixed relative to external contexts (comparison classes).

[23]I use 'non-traditional' here rather than 'non-classical' since, as we shall see, on most of the big semantic issues, Graff Fara remains classical.

height than is typical depends on what standard is in play. If that standard is basketball players or Dutch men (who are, on average, loftier than most), I am not greater in height than is typical. I might be greater in height than is typical if the standard in play is British men, but I may still not count as tall if 6ft 2in doesn't count as significantly taller than that standard.[24]

As well as the norm relativity and interest relativity of terms, Graff Fara introduces four constraints on the meaning of terms. The first three of these, she takes to be undisputed: *The clear case constraint:* For every term, there are cases that that term would apply to independent of context.[25] *Relational Constraints:* Some predicates have relational features which hold between cases. This constraint effectively describes the penumbral connections appealed to by supervaluationists (see §1.6.1 below). Another way to think of these is as akin to monotonicity. Imagine a number of sets, each of which is a proper superset of the previous one. Some predicates that apply to members of those sets will be upward monotone, others downward. For example, if the sets contain individuals under a specific height, 'tall' would be upward monotonic and 'short' would be downward monotonic. *Coordinated Standards:* This constraint covers antonyms (and presumably linguistic negations), such that anything that is, say, tall, is excluded from counting as short or not tall.

On top of the above three constraints is the *similarity constraint:* "Whatever standard is in use for a vague expression, anything that is saliently similar, in the relevant respect, to something that meets the standard itself meets the standard; anything saliently similar to something that fails to meet the standard itself fails to meet the standard." (Graff Fara 2000, p. 57).

Graff Fara takes the similarity constraint to be less obvious and more in need of justification. At the same time, it is also what lies at the heart of her account of vagueness (i.e. her treatment of the Sorites paradox). The major move she makes is to argue that the similarity constraint falls out of

[24]This reading of Graff Fara seems justified on the basis of the kinds of examples she uses. For instance:

> "'John is tall' is to be analyzed as meaning ... 'John has significantly more height than is typical'... 'Mickey is old for a mouse' ... as meaning 'Mickey has significantly more age than is typical for a mouse'." (Graff Fara 2000, p. 65)

However, later on that page, the line is blurred slightly when we are told:

> "Sometimes 'a lot' can mean 'significantly more than is typical', but at other times it can mean 'significantly more than is is wanted or needed'. And also it can mean 'significantly more than is expected'.

[25]We will see a reason to doubt this constraint in Chapter 2.

an acceptance of the far more intuitive notion of interest relativity.

The key idea is that two things (heaps, coloured cards, people with respect to their heights etc.) can be qualitatively distinct but nonetheless count as the same for present purposes. Sameness is interest relative. Whether or not something is in the extension of a predicate can follow from whether or not something that is, in the relevant respects, the same, itself falls into the extension of that predicate. So if a link can be made between being the same for present purposes and being saliently similar, the similarity constraint will fall out of interest relativity. Graff Fara forges this link by analysing interest relative sameness in terms of costs and benefits. Her example is coffee. She can, presumably truly, say that she uses the same amount of coffee each morning to make her daily dose. The same measure gives, for the purposes of making a drink which will have the same effect on you, the same effect. A few grains don't make a difference. In fact, for those purposes, to worry about sameness of grains would impose a higher cost than it would a benefit. This is not always the case. The pharmacist of old who is making up a tincture might need to care a great deal about the number of grains of a drug it contains. The benefit of not giving a patient an overdose far outweighs the cost of taking care in the measurement of a drug. So interest relative sameness is cached out in terms of a cost benefit analysis of whether or not to discriminate a difference. Now Graff Fara can link interest-relativity with salient similarity, since, she argues, there cannot be a cost associated with discriminating two things when one or both of them is not even a live option, and the notion of being a live option is supposed to be a euphemism for salience.

Salient similarity is used by Graff Fara to solve two of three questions that she associates with the Sorites paradox:

Graff Fara's three questions:

1. Semantic question: The denial of the universal Sorites premise or any instance of an induction step entails a sharp boundary. One of the following must be answered:
   (a) How can sharp boundaries be compatible with the vagueness of predicates?
   (b) If we deny the Sorites premise and sharp boundaries, how is this compatible with classical logic?
2. Epistemological question: If there is a sharp boundary, how can we not know where the boundary is?

3. Psychological question: We need to deny the key Sorites premise, but why do we find it so plausible?

Graff Fara defends bivalence and wishes to keep a classical logic. She takes this to entail sharp boundaries for vague predicates. That places most explanatory weight on an answer to the psychological and epistemological questions.[26] The paradox is analysed in terms of sameness. The Shard is tall for a a building in London, Granny's bungalow isn't. If the two are salient, there may be many purposes relative to which we may discriminate between the tallness of the two. That means, for Graff Fara, there is a switch from tall to not tall somewhere between the two. Where this boundary is, on Graff Fara's account, changes with our interests. When we come to compare any two buildings of similar but non-identical height, the cost of discriminating between the two will be high.[27] This means that, wherever we look, even had the sharp boundary been there before, it would slip away from our mental grasp as soon as a case on either side of the divide became salient. The epistemological question is thus answered and the psychological plausibility of the Sorites premise arises because every instance (either of the universal or inductive premise) becomes, on considering it, eminently plausible.

So Graff Fara's account seems, just like Raffman's, to be a solidly classical semantic picture with some contextual shift, psychology and epistemology on top. However, the more interesting proposal in her paper comes right at the end in her account of vague adjectives. Graff Fara takes comparative constructions to be instructive of how to interpret vague adjectives in non-comparative constructs. $\texttt{tall}(x)$ will be interpreted as some kind of numerical value which stands for some degree of how tall x is. This is not meant to be a degree of truth (see §1.7 below). The value just represents a distance from a central case such that the greater the distance, the lower the value. For comparatives, the morpheme construction '-er (than)' is taken to introduce the semantic structure for the sentence in which it is a part:

$$-\text{er} = \lambda G.\lambda y.\lambda x.(G(x) > G(y))$$

[26]Graff Fara maintains that accounts that are similar to hers in spirit needn't accept bivalence. As she states, a sympathetic supervaluationist could endorse the salient similarity constraint with respect to precisifications (see below).

[27]Graff Fara isn't explicit here, but I think that the cost metaphor is supposed to relate to something like the offence to our sensibilities (or maybe competence), of drawing a sharp boundary for a vague predicate.

In other words, -er constructions are only true when the degree to which the predicate applies to the object in the subject NP object is greater than the degree to which it applies to the object in the complement NP. Graff Fara then supposes that all vague adjectives have a slot for a morpheme that is filled by a silent morpheme in pronominal constructions like 'Mike is tall'. The silent, *absolute* morpheme is modelled on Kennedy (1997):

$$\varnothing_{ABS} = \lambda G.\lambda P.\lambda x.(G(x) \mathrel{!\!>} (NORM(G))(P))$$

This morpheme provides a comparable semantic structure to '-er' except that '!>' means saliently greater than, and '$(NORM(G))(P)$' returns a value for whatever the norm for being G is with respect to being some other way, P. P is supposed to be filled in either by a 'for a' clause (The Shard is tall for a London building) or from some contextually provided comparison class.

The substantial move that is being made here is twofold. First, what might standardly be taken to be a pragmatic aspect of vague predicate use is being written directly into the semantics of such predicates via '!>' and '*NORM*'. Second, the meaning of vague predicates such as 'is tall' is no longer truth-functional, but degree functional. The value of '$\texttt{tall}(x)$' will be the result of applying the measure function for `tall` to x. This value will represent the degree to which x is tall. What this amounts to is the introduction of a new basic type into the semantics, say type d for distance.[28] Keeping other aspects of the semantics simple, this would make `tall`, and presumably all other vague predicates be of type $\langle e, d\rangle$. All predicative uses of vague terms (or at least adjectives) will be evaluated as a comparison of values of type d.

At this point, our goal is not to raise objections, but a few comments can be made, some of which we will return to later. The distance measure for applying an object to an expression such as '(is) tall' will be specific to '(is) tall'. It would be senseless to compare an '(is) tall' measure with an '(is) blue' measure. This might lead one to think that Graff Fara is defending a non-reductive account of the meaning of '(is) tall'. One which itself uses 'tall'. For example, since $\texttt{tall}(x)$ is interpreted as some degree or distance marker, it could be interpreted as: *to degree d, how close to a clear case of being tall x is*. However, Graff Fara informally spells out the meaning of 'is tall' as "a function from individuals to degrees on a height scale". We

[28]See (Kennedy 2007, §2) for discussion and references.

might then ask, how we should translate '(is) tall' on the formal account into heights. Maybe, using the approximation we are provided with (Graff Fara 2000, p. 73) "has d amount of height".

'Tall' has to do with heights and so 'x is tall' will be interpreted as *x has a significantly greater amount of height than is normal for a P*. Notice, however, that this assumes that the degree value for height can be given context independently: There is, context independently, an amount of height that x has, and some amount of height that Ps normally have. For an expression such as 'is tall' the above account may work, since the base line for how much height something has will always be 0. If tallness is always about height, there may be a context-independent minimum for height, namely no height. Hence every object could be assigned a height value and these could be compared with the norms for height values to give the semantics of 'tall'. But what about when we have a predicate like '(is) green', let alone when we think of more abstract adjectives like 'brave' or 'honest'?

Maybe an example will help to bring out the point. Say we fix a particular shade, or range of shades as the baseline for 'green'. All objects within that shade range would receive the maximum amount of greenness on the scale. The farther something's shade is away from the range of baseline shades, the less greenness on the scale it gets. Now, some x is green if the amount of greenness it has is greater than some norm for some contextually determined comparison class. Given that x is some shade, it will always have the same degree of greenness independently of the class it is compared to. The question we should ask is whether we should ever expect that degree of greenness to change from context to context. A reason for thinking it might is that what counts as a maximum in one context might not count as a maximum in another. The Queen's lawn might have the shade that would make it a clear case of green. That means that the Queen's lawn must be at least as green as every other possible comparison class. But that means that the standard for 'green', against which the Queen's lawn and other comparison classes are measured, is fixed (this actually follows merely from the clear case constraint). There is a tension here. Why should we expect whether or not something is green to vary from case to case, but whether or not something is a clear case of green to remain fixed? There is a parallel here to higher-order vagueness. Accounts such as supervaluationism and fuzzy semantics (see below) have been criticised for drawing sharp boundaries at the second order of vagueness (between things that are definitely G and not definitely G). Graff Fara makes a good case that our interests determine our standards for when something is G, but at the same time implicitly commits herself to the view that no contexts or interests can shift out standards for when things are or are not definitely G.

All of the context/interest relativity of Graff Fara's model comes built into '!>' and '*NORM*' which are used to evaluate the value that the object applied to the predicate receives. So, by its own lights, it may be problematic for a contextualist account to embrace a context-independent notion of a clear case.

1.6 Vagueness in the Literature – Ambiguity*

All of the approaches considered so far hold that, relative to a context, utterances made using sentences containing vague terms express/are associated with a single Boolean proposition. The 'single' part of this assumption is dropped by theories we will now consider. These theories (daughters of node [10]) divide into two camps. Either there is indeterminacy in which proposition, of multiple propositions, a vague utterance expresses (daughters of [101]). Or multiple propositions are not expressed/part of the meaning of the utterance, but are, in a sense to be clarified, *admissible* or acceptable extensions/interpretations of the utterance (daughters of [100]). However, this distinction is subtle and details vary from author to author.

Examples of [101] daughters are Nihilists (Braun and Sider 2007) (B&S)[29] and plurivaluationists.[30] Both emphasise indeterminacy over which proposition is being expressed. The difference is that plurivaluationists think that what we say can be true (false) whereas nihilists deny this.

Examples of [100] daughters are supervaluationists (Kamp 1975), (Fine 1975), (Keefe 2000), and a version of probabilistic epistemic semantics laid out in (Lassiter 2011) and (Frazee and Beaver 2010). For supervaluationists, the propositions expressed by vague utterances are vague. However, these meanings can be modelled via classical *sharpenings* or *precisifications* of these vague expressions. Lassiter's probabilistic epistemic account has just one proposition expressed from multiple allowable precise ones. However, one cannot be sure which one's interlocutor is expressing. Vagueness is captured as a probability distribution over precise propositions. Lassiter's

[29] B&S cite Frege as an original nihilist in virtue of the view Frege held, that vague sentences cannot have a truth value. Frege doesn't make the cut on this division of the literature since there is, to my knowledge, no evidence that Frege associated vagueness with the expression of multiple thoughts/vague terms having more than one sense.

[30] A discussion on plurivaluationism can be found in (Smith 2008). He does not defend a simple plurivaluationist account, but rather a fuzzy version of it. See §1.7.

and Frazee and Beaver's views (node [1000]) are really outliers in this group. They can also be seen as connecting with other epistemic views (node [1111]). The difference is one of emphasis. Lassiter, and Frazee and Beaver emphasise that we can approximate to the proposition being expressed. Epistemicist views emphasise the unknowability of the right meaning without directly appealing to other propositions.

It should be apparent that the difference between multiple and single proposition views is not that great. Epistemicists emphasise the unknowability of the proposition expressed. Supervaluationists emphasise that multiple propositions are admissible as sharp meanings for what was vaguely expressed. Plurivaluationists and nihilists emphasise that it is indeterminate which of many propositions are expressed.

I will now discuss some general arguments that motivate the consideration of multiple propositions. I will then discuss some specific accounts.

Argument from Analogy

A fairly common thread in multiple proposition views is to draw an analogy with ambiguity.[31] The word 'green' is ambiguous and therefore, so is:

(a) Janet decided to buy a green car.

One thing (a) could (be used to) mean is that Janet decided to buy a green coloured car. Alternatively, she might have decided to buy an environmentally friendly car.[32] Views on ambiguity are pretty wide and varied, but within philosophy, a common approach is to assume that the sentence has two or more propositions for its meaning, but that an assertion of the sentence expresses, in virtue of the intentions of the speaker, a single proposition. Speakers' intentions disambiguate ambiguous sentences. Independent of an assertion or utterance, it just doesn't make sense to ask whether an ambiguous sentence is true or false *simpliciter*. As commented on in §1.1, on many peoples' views, it never makes sense to even ask if a non-ambiguous sentence, independent of an utterance of it, is true or false. However, letting this pass, one might think that, without the appropriate speaker's intention, an utterance of 'Janet decided to buy a green car' actually expresses two propositions, namely, the two disambiguations of the sentence.[33]

[31] This is not shared by either (Lassiter 2011) or (Frazee and Beaver 2010).

[32] It should be emphasised that there are many more possible disambiguations of this sentence.

[33] Fine does not discuss the utterance/sentence distinction. He talks only in terms of sentences, which he takes to express all of their disambiguations.

It is at this point that the analogy between ambiguity and vagueness is supposed to kick in. The problems of the Sorites paradox and of giving a semantics for vague terms are proposed to be solvable by assigning a richer meaning to vague sentences and utterances of them. Utterances which use vague terms are just like utterances which use ambiguous terms, but which lack a disambiguating intention. Such ambiguous utterances express multiple propositions. The line of thought is that vague utterances behave similarly. This is meant to be an analogy. Vagueness is not being argued to be ambiguity, but rather somewhat like ambiguity.

Let us say that vague expressions are ambiguous* rather than ambiguous in the traditional sense. The difference can be seen by looking at other cases of ambiguity. For example, vagueness is certainly not like a river bank/financial bank kind of ambiguity. With this sort of ambiguity, the same word just happens to mean two different things.[34] Nor is vagueness the same as the kind of ambiguity one finds in different, but related meanings of the same word. For example, there are fourteen classes of disambiguations in the OED for 'green' as an adjective, hence there are probably at least that many ways that 'Janet decided to buy a green car' can be ambiguous. However, for any one of these ways, assuming that all are vague[35], there will be more than one proposition expressed. For every disambiguation, there are multiple disambiguations* and disambiguations* are not the kind of things you can find being described in a dictionary.[36] Any proponent of the ambiguity* view therefore has an undischarged duty to spell out what kinds of propositions are expressed by (uses of) sentences like (a).

The Argument from Properties

As well as an analogy to ambiguity, the multiple proposition view is also argued to follow from a particular kind of view about properties in the world (Braun and Sider 2007, p. 134). This argument presupposes that vagueness

[34]Evidence for this kind of ambiguity is, for example, that other languages have different words for the disambigated senses.

[35]It is unclear whether one of these is vague. In the OED, entry fourteen for green is: 'Particle Physics. Of a quark: having the colour green'. I am advised that there are no borderline cases of being a green quark. Thanks to Alistair Butcher for his advice on this matter. If I have not got this quite right, the fault is entirely my own. An example not in the OED was also raised by Robin Cooper, that of a green cut in Prolog programming. This, too, does not seem to admit of borderline cases.

[36]I do not wish to labour this or to make a methodologically loaded ordinary-language point. I only wish to mark that it might be a mistake to take the ambiguity analogy too seriously.

is not a worldly phenomenon. Its first premise is that properties in the world are sharp and that different objects in the world, all of which we could call 'green' have different colour properties. Second, predicates of natural languages are referring terms and their referents are properties. Third, vague predicates cannot be said to refer to a single property since we call objects with different colour properties 'green' (from the first premise). Finally, all candidate properties are equally good, therefore no single property can be said to be the one referred to over and above any other. Any sentence containing a vague term therefore has multiple equally good candidates that could count as its meaning.

The first thing to notice about this argument is that it doesn't quite reach the same conclusion as the ambiguity analogy. The ambiguity* view holds either that sentences containing vague terms or utterances of such sentences express multiple propositions (just like ambiguous sentences are supposed to do).[37] The argument from properties doesn't get that far. It's conclusion is only that the candidate meanings are all as good as each other. Nothing is said about what is expressed.

Second, the 'equally' in 'equally good candidates' seems to miss something key to vagueness. One example of a colour property might be the property an object has when it looks yellowish-green (in normal lighting conditions or whatever). The quandary is whether we can refer to such a property using 'green' or 'yellow', or whether we cannot do so with either. Compare this to the colour-property/properties the leaves of healthy looking lawn-grass have. Surely, if we're talking properties which are candidates for the referent of 'green', the latter are far better than the former.

I will return to the notion of properties and equally good candidates shortly where it will turn out that the kinds of properties that are being appealed to by ambiguity* views are not the kinds of things that one might normally consider to be properties, but rather classes of properties. First, I will give a brief outline of supervaluationism and nihilism. This will include discussion on what kinds of propositions we should assign as the meanings of (uses of) sentences containing vague terms. I will then highlight the key differences between the two positions.

[37]For an alternative conception of the multiple proposition view where this commitment is not made, see §1.6.1.

1.6.1 Supervaluationism versus Nihilism

Kit Fine starts out with a problem for accounts of vagueness that posit I (for indeterminate) as a third truth value. The motivation for such accounts comes from examples of genuinely irresolvable cases of whether a vague term applies to a borderline case. Fine points out that while a third truth value seems like a natural way out of the puzzle, it creates far more problems than it solves. Principally, such problems relate to *truths on a penumbra*. Penumbral truths arise from logical connections that hold between statements that are indeterminate with respect to truth value. Many of the examples given for these turn on theorems and contradictions in sentential logic. For example, if A is indeterminate and B is indeterminate, $(A \vee \neg B)$, $(A \wedge \neg B)$, $(A \rightarrow B)$, $(A \equiv B)$ would all seem to be indeterminate, too.[38] However the following should all still be true: $(A \vee \neg A)$, $(A \rightarrow A)$, $(A \equiv A)$. And $(A \wedge \neg A)$ should be false. Fine demonstrated that there are no truth-functions for logical connectives that can respect these patterns. In other words, inclusion of a third truth value that respects intuitions over theorems and contradictions breaks the truth-functionality of connectives.

Fine's suggestion was to view vague sentences as imprecise. Associated with every sentence would be a set of admissible precisifications. Certain conditions are placed on how sentences can be made more precise. For example, a true (false) sentence would never change truth value under precisification, however, a sentence that was neither true nor false could become true (false) under precisification. Truth gaps are eliminated by precisification. Also, for every sentence, be it true, false, or neither true nor false, there is a precisification under which it is either true or false. Precisifications can be understood as propositions with Boolean values. Hence, vagueness can be viewed as imprecision where all of the ways to make the imprecision precise are analogous to all the ways one can disambiguate an ambiguous sentence. Unlike ambiguity and disambiguations, truth gaps and precisifications are more structured. Crucially for Fine, the precisifications for vague terms respect penumbral truths. For example, if a sentence is neither true nor false, then there will be a level of precisification for that sentence such that every precisification will be true or false. Under any precisification $(A \vee \neg A)$, $(A \rightarrow A)$, and $(A \equiv A)$ will be true and $(A \wedge \neg A)$ will be false. Since indeterminate is not a truth value, sentences of the form

[38] Whether we take a maximising strategy or a minimising strategy with respect to indeterminacy, if both components of a binary connective formula are indeterminate, the whole will be indeterminate.

$(A \vee \neg B)$, $(A \wedge \neg B)$, $(A \rightarrow B)$, $(A \equiv B)$ may be neither true nor false, but all of them will have precisifications under which they are either true or false. Penumbral connections have a special role to play in the supervaluationist account of Sorites arguments. I return to this point in §1.6.2.

What makes supervaluations super is that the class of admissible precisifications for an imprecise sentence, each of which has a Boolean value, can be used to give a valuation for the imprecise sentence. Either all precisifications are true, all are false, or some are true while some are false. Imprecise sentences can be given values of (super)true and (super)false in the first instances, and neither (super)true nor (super)false in the last. "Truth is supertruth, truth from above" (Fine 1975, p. 273).

Fine talks only of sentences, but utterances and context-sensitivity can be built in to this picture, too. For example, we might say that the class of admissible precisifications for an utterance changes with context. If the comparison class is professional basketball players, then different precisifications of 'tall' will be admissible than if the comparison class is primary school pupils. It might be supertrue that I am tall on the latter, but superfalse that I am tall on the former.

As Braun and Sider readily concede, their form of Nihilism is structurally very similar to supervaluationism. Utterances are associated with multiple propositions, and propositions have Boolean truth values. This will be discussed below. Sentences are imprecise, and it is rare for an utterance that uses a vague term to be precise to the extent that only a single proposition is expressed. Rather than talking of precisifications B&S use the term 'legitimate disambiguations'. The distinction between the views comes down to what can be a bearer of truth.

Supervaluationism effectively employs two concepts of truth. One applies only to propositions/precisifications. The other is supertruth. Supertruth is defined in terms of the truth values of the propositions/precisifications for the sentence/utterance in question. If we were to accept that all natural language terms are, to some extent, vague, then truth for propositions/precisifications would only really be a formal concept that never applied to natural language expressions. At the very least, given that most statements made in natural language are vague, truth for precisifications can rarely apply to statements in natural language. When we judge sentences/utterances to be true (false), if supervaluationism is correct, we are really judging them to be supertrue (superfalse). B&S's nihilism has two phrases using 'true', but strictly employs only one concept of truth. Nihilism denies that utterances using natural languages can be true (false) because propositions are truth-bearers and utterances in natural languages "rarely if ever" express a single proposition. However, just as all precisifications can be true, all legitimate disambiguations for an utterance can be true too. If they are all true, then the utterance is *approximately true*. Approximate truth is not truth, nor is it truth-likeness (Niiniluoto 1977). Approximate truth is, in a sense, absentminded pretend truth. If we are not paying attention to the vagueness of the language, we absentmindedly ACT AS IF it were precise (Braun and Sider 2007, pp. 135-7). Acknowledging vagueness destroys the illusion. Doing philosophy, it seems, can lead us to realise that we were merely treating what people said as if it were true (false) when really, all this while, it failed to be either.

Despite this difference over what kinds of things can be true, the two positions are remarkably close when it comes to their treatments of meaning. Their explanations of the semantics going on in Sorites arguments are roughly equivalent, even if the psychological aspects of the explanation are not. Universal Sorites premises are superfalse/approximately false because they are false on every admissible precisification/legitimate disambiguation. Nonetheless, there will be a number of cases in every Sorites for which applying the vague predicate in question is neither supertrue/approximately true nor superfalse/approximately false. In these cases, speakers may be unsure of whether to apply the predicate or not.

Supervaluationism without Ambiguity

The above characterisation of supervaluationism has been rejected by some supervaluationists who claim that they do not view vagueness as similar in any way to ambiguity and would reject any talk of there being (sharp) propositions expressed by uses of vague terms at all. Rosanna Keefe (2000), who is fairly faithful to Fine's formal characterisation of supervaluationism, nonetheless rejects Fine's description of vagueness as "ambiguity on a vast and systematic scale". Keefe suggests that supervaluationism was never meant to be a theory about what the meanings of vague terms are, but just about whether sentences using vague terms are true or false. Understood this way, supervaluationism can stay silent on whether some vague proposition is expressed or on whether some number of sharp propositions are. Indeed, as Keefe insists, supervaluationism can be stated independently of talk of propositions at all.

Supervaluationism, on Keefe's conception, is the view that precisifications are simply ways that vague utterances/sentences COULD be made more precise:[39]

> "ambiguous expressions have several ACTUAL meanings, whereas when there is semantic indecision between some precise extensions, those extensions reflect, at best, HYPOTHETICAL (sharp) meanings: the actual meaning is vague and typically univocal". (Keefe 2000, p. 157), her emphasis.

This doesn't suggest that Keefe adopts the view that there is one proposition expressed by a vague utterance, even though this proposition is vague:

> "We may, for example, resist the inference from 'linguistic expression X has a meaning' to 'there is some item which is X's meaning'. Or 'X's meaning' may indeterminently refer to each of the relevant range of precisifications". (Keefe 2000, p. 158).

Refusing to take there to be some thing that is itself THE meaning of an utterance is a plausible idea which will be returned to when the positive account is developed.

Keefe goes on to say more about meaning in a response to (Fodor and Lepore 1996)'s objection that precisifications, as sharp entities, violate our everyday understanding of what vagueness is:

> "But this constitutes no objection to the theory, for the claim is that it is the quantification over all precisifications that capture the meaning of the natural language predicates; the individual precisifications need not." (Keefe 2000, p. 190).

At this point there is an air of splitting hairs about drawing a distinction between Keefe's position and one like Fine's. Keefe may not wish to commit herself to some view about what kind of proposition(s) can be expressed by uses of vague terms. However, the meaning of sentences containing such terms will still respect a long standing idea about meanings that goes back to Frege. Frege describes thoughts as being "something for which the question of truth arises" (Frege 1918/1956, p. 292). Keefe's meanings, characterised by quantification over precisifications, are supposed to do just that, albeit

[39]This is not the same as 'underspecification' (Bach 1994). Precisifying an utterance is a way of caching out what that utterance anyway says (albeit vaguely). Underspecification (also, 'underdetermination') concerns taking what is encoded by a sentence and filling it out to get a full utterance meaning.

for supertruth. The differentiation between the two positions, then, becomes whether we wish to say that the meaning of an utterance is characterised by the propositions it expresses, over which its truth is defined, or, whether the meaning of the utterance is characterised as a quantification over a range of propositions (that may not actually be expressed), that determine its truth. For most purposes, this may well be a distinction without a difference.

Plurivaluationism

There is an account of vagueness that looks a lot like supervaluationism, described as *plurivaluationism* by Nicholas Smith (2008). Smith points out that some positions that are described as supervaluationist are pretty different in terms of structure. Supervaluationism involves two different kinds of semantic models, one of which is non-classical. The non-classical model operates on the level of supertruth and defines the truth values of complex formulas non-recursively. For example, disjunctions which are supertrue may have one, two or no supertrue disjuncts. The precisifications can then be seen as admissible classical extensions of the non-classical model.

Plurivaluationism does not contain a non-classical model. It posits a range of classical models for every utterance, but no one of these is the uniquely intended interpretation. If all of the models/interpretations are true, then it does not matter which of the models/interpretations we take, and so the sentence is true simpliciter. If not, then it is not.

Spelled out this way, Braun and Sider's view comes out as a nihilistic form of plurivaluationism, rather than a nihilistic supervaluationism. Keefe's view is completely supervaluationist. Somewhat surprisingly, although Fine's view is supervaluationist in implementation, it is a little plurivaluationist in spirit (in the sense that vague utterances are really supposed to mean a collection of their precisifications). I will return to plurivaluationism again in §1.7 to discuss Smith's own version of it, *fuzzy plurivaluationism*.

The kinds of propositions that sentences containing vague terms are modelled as expressing/having as their precisifications are, in part, motivated by certain penumbral connections. To end this section, I will address what kinds these are.

1.6.2 Equally good candidates, properties, and clusters of properties

Earlier we saw an argument that started with the premises that properties are sharp and that predicates refer to properties. The conclusion of this argument was that vague predicates each have a number of equally good candidates to count as their meanings. The further conclusion (that wasn't quite reached) was that all legitimate/admissible good candidates are expressed. I do not wish to get into the (meta)physics of determining what properties (there) are, but I do want to say something about the kinds of properties that the ambiguity* view commits itself to (assuming that predicates refer to/pick out properties).

A more traditional theorist who defends a single proposition view might think about properties, too. They might realise that 'green' can be used to refer to objects with any of many colour properties. After all, green things can have any of many different shades and some of them are quite dissimilar. Nonetheless, the traditionalist doesn't see this as a threat to the view that, relative to a context, utterances of declarative sentences containing 'green' refer to a single proposition. 'Jen's cabbages are green' can be true if Jen has a dark lavish savoy or a lustrous spring variety. If savoy cabbages and spring cabbages have different colour properties, whatever 'Jen's cabbages are green' expresses, the idea is that it can be true of things with either of those properties provided they are cabbages that are Jen's.

On the single proposition view, whatever 'green' is supposed to refer to, it cannot be meant to be something as fine-grained as the colour properties that some green cabbages have and that others don't. Instead, it must refer to something broader. A natural thought is that 'green' expresses either a cluster, or a range of properties. On a traditional view this range is modelled as some set, and has sharp boundaries. Within the set, objects may have any of many (fine grained) colour properties. Traditionalists accept a multiplicity of properties as being referred to by vague terms without accepting the need for multiple propositions.

The idea for multiple proposition accounts was that in borderline, unclear cases, some equally good candidates for 'green' apply while others do not.[40] This suggests that what multiple proposition theorists mean is that there

[40] If the good candidates were highly fine grained colour properties, this idea would not be unique to borderline cases. All good candidates for 'green' would apply to some but not all cases of green things even if they were clear cases of green things. I take this to be a further reason against the possibility that it is those kinds of properties being talked about here.

are different ranges of properties that are being referred to. So, for colour terms like 'green', we might think of precisifications/disambiguations as forming a set of concentric circles. The innermost circle would contain all the colour properties that only, say, the definite cases of green have. The farther out we progress, all the former properties are included, but new ones are added. Eventually, colour properties are added that, were one to see an object with that property, one would find it hard to judge whether it was green or not.

There is still something odd about that idea. In what sense is the outermost ring (the broader range of properties) as good as ranges closer in? One possibility is that there might not be a good enough reason to include one but not the other.[41] Yet there is a natural sense in which the properties the clearer cases have are better candidates for what 'green' refers to than the less clear cases. Supervaluationists and Nihilists attempt to capture this with supertruth and approximate truth; in Fine's words, "truth is secured if it does not turn upon what one means." (Fine 1975, p. 278).

To end this section, I would like to mark the slightly peculiar journey one must go on to get to this point (either as a single or as a multiple proposition theorist). The journey starts with the plausible idea that expressions like 'green' have something to do with the colour properties that things have. Considerations of use lead us to realise that things with many different colour properties are called 'green'. Consideration of Sorites arguments force us to think about which of these properties something that is green can truly have. This forces us away from the intuitive idea that terms like 'green' relate to colour properties, and towards a new idea that the meaning of 'green' should be given in terms of a collection of colour properties. Some people assert that there is exactly one collection of properties that things that are truly called 'green' have. Others deny that there is any principled way of choosing this collection.

It is not the task of this chapter to weigh up which, if either, of these positions is more plausible. However, maybe placing too much weight on the Sorites paradox itself is creating the problem. The paradox presents a puzzle for what the true applications of vague predicates are. If the goal of studying vagueness in philosophy is to find a solution to this puzzle, it is hardly surprising that multiple incompatible accounts will all appear to do so – there will be different inferences to the best explanation dependent on what assumptions one starts with. It is therefore also hardly surprising that there

[41] This is akin to saying that vague terms are tolerant in the sense of (Wright 1975).

is no clear way to choose between these competing explanations since it is highly likely that the assumptions that lead to them differ fundamentally. The suspicion (§1.1) is that paradoxes like the Sorites should be available as a test of an independently grounded and motivated theory of meaning, not as part of the grounds for such a theory.

1.7 Vagueness in the Literature – Degree Theories

A plausible idea was mentioned at the end of the last section: Expressions like 'green' have something to do with the colour properties that things have. However, the colours of some things are hard to define. We have devices in natural languages to cope with phenomena like this. For example, in English, adjectives can be suffixed with '-ish' to indicate that something is not exactly some way or another.[42] Add to this that, as well as being more or less green, something can be more or less green than something else, and another natural sounding idea presents itself: Perhaps the meanings of such terms should be represented in terms of degrees.

As we saw in §1.5.2, incorporating a notion of degree into semantics can be consistent with an entirely classical approach. Hence, nothing in the previous paragraph should be objectionable to any vagueness theorists. Controversy begins when we try to say exactly how and why degrees matter.

Degree theories have been saved until last in this whistle-stop tour of the literature because some of them signify the most radical departure from traditional ideas of how to represent meaning. However, many degree theories (including Graff Fara's, Lassiter's, and Frazee and Beaver's) are not so radical. The literature on degrees is itself large. In this section I will mostly discuss *pure degree theories*. With the possible exception of the [110] interpretation of verities, pure degree theories are found as [0] daughters on the theory tree in §1.3. Pure degree theories replace the Boolean values of 0 and 1 with the set of reals in the interval $[0, 1]$. The differences between pure degree theories are in what these values indicate (degrees of truth, degrees of closeness to clear truth, or degrees of uncertainty) and which logic should govern them (fuzzy logic or probability logic).

The best known pure degree theory is fuzzy semantics (Zadeh 1965), (Goguen 1969). Fuzzy semantics was developed in the 60s and has seen wide-spread application in fields such as engineering. It has come under

[42] Another example is the Finnish suffix '-hko/-hkö'.

considerable fire as applied to vagueness in philosophy.[43] A vigourous defence of fuzzy semantics has been made recently by Nicholas Smith (2008). I will consider some of his arguments below.

The other kind of pure degree theory also has its roots in the 60s, albeit 200 years earlier. Bayes' ideas for a probability calculus, later built upon by Kolmogorov (1950) were first applied to vagueness by Dorothy Edgington (Edgington 1992), (Edgington 1997). The simple and elegant suggestion made by Edgington was that the tried and tested mechanics of probability calculus that had already been put to use in philosophy for modelling credence could also be employed to model what Edgington calls *verities*. As mentioned in §1.3, there are different possible interpretations that can be given to verities. Edgington understands verities as closeness to clear truth. However, her main concern ((Edgington 1992) ,(Edgington 1997)) is to develop a logic for vagueness, not a semantics. That is why there are three possible spaces for verities on the semantics decision tree in §1.3: ([110],[01],[00]). The view I will come to defend will be the [00] reading of verities. Below, in §1.7.2, I will simply sketch the formal details of the verities account. I will give a detailed discussion of the interpretation of verities in chapter 8.

Very recently another kind of pure degree theory has been developed. This approach is also Bayesian, but meaning representations are taken to be straightforwardly probabilistic. Two versions are in manuscript form. One pursues a top-down approach where meanings are distributed over worlds (van Eijck and Lappin 2012). The other, drafts of which have only just been written, develops a more bottom up approach based on a probabilistic representation of speakers' judgements (Cooper et al. 2013).[44]

As well as pure degree theories the literature also contains a variety of what I shall call *degree augmented theories*. Degree augmented theories are exclusively found as [1] daughters in the theory tree (§1.3). These theories take an existing position in the literature and incorporate elements of one of the pure degree theories. For example, degree supervaluationism (Williams 2011), fuzzy plurivaluationism (Smith 2008), and bayesian epistemic semantics (Lassiter 2011). Not all of these involve degrees of truth. Bayesian augmented approaches tend to employ probability calculus in some respect, but retain the notion of truth native to the approach being augmented. For example, Lassiter's probabilistic epistemic logic leaves truth as it is in a

[43]Strong examples include (Kamp 1975), (Williamson 1994), and (Edgington 1997).

[44]Skyrms (2010) has also recently developed a probabilistic semantics. He does not apply it to vagueness however.

standard classical modal logic.

Below, I will outline some of the basic motivations and semantic commitments of pure fuzzy semantics and pure Bayesian theories. The position that I will eventually be arguing for will be a probabilistic pure degree theory and so I leave discussion of other such theories until after my own view has been put forward. Finally, I shall give a brief overview of some of the degree-augmented theories. Again, some of these will be returned to later once my own account has been put forward.

Before moving on, a small hiatus with respect to information is in order. A category of theories on vagueness that has not been discussed is *information theoretic* theories. Quite a few of the degree theories also count as information theoretic especially (Lassiter 2011), (Frazee and Beaver 2010), (van Eijck and Lappin 2012), and (Cooper et al. 2013). However, there are information theoretic approaches that are not degree based (especially (Floridi 2011)). Information theoretic approaches have their historical origins in the mathematical theory of communication proposed by Claude Shannon (Shannon 1948). However, their close recent ancestor is Dretske's (Dretske 1981) notion of semantic information. The account of vagueness that will be developed over the course of this thesis, as well as classifying as a degree theory, will also count as being information theoretic. I therefore save detailed discussion of similar accounts until chapter 7.

1.7.1 Fuzzy Theories

The first fuzzy semantic treatment of vagueness was, presented in Goguen (1969) where it was suggested that meanings of vague terms can be represented by fuzzy sets. Classical sets can be represented by functions into Boolean values. Objects that map to 0 are not members of the set, objects that map to 1 are. Fuzzy sets are functions into $[0, 1]$, so objects can be members of the set to various degrees. The intuitive idea is that fresh healthy grass might be something in the fuzzy set to represent the meaning of 'green' and so would get a value of 1. Under-watered grass that is beginning to die might only be greenish. If it is as green as it is yellow, perhaps it would get a value of around 0.5.

Just as in classical set theory and semantics, in fuzzy semantics, intersection, union and complement can be used to define the logical constants for conjunction, disjunction and negation. Let $[q]$ be the fuzzy valuation of the proposition q:

$$\begin{aligned} [\neg q] &= 1 - [q] \\ [q \wedge r] &= min\{[q], [r]\} \\ [q \vee r] &= max\{[q], [r]\} \end{aligned}$$

At the limit fuzzy conjunction, disjunction and negation is classical. Things change radically for values between 0 and 1 which has tended to be the source of objections to the system applied to natural language.

A more general concern with the fuzzy system that has been raised is what the values between 0 and 1 mean. Some in the literature seem to feel discontented with what they represent (Graff Fara 2000), (Williamson 1994). Goguen himself is fairly noncommittal. His paper was entitled 'The logic of inexact concepts', but Goguen tells us that 'concept' should be taken as a metaphorical device for talking about meaning. However, given fuzzy logic as a framework with which to model natural language semantics, it is left fairly open how we should interpret the notion of a fuzzy set.

Smith, who defends a plurivaluationist version of fuzzy semantics (Smith 2008) argues that fuzzy semantics and worldly vagueness go hand in hand. If one thinks that the world contains vague properties, one may be tempted to see fuzzy sets as representing properties in the world. That way, it would be indeterminate whether some object, a, has some property *redness* and the truth values of sentences like 'a is red' would be isomorphic with the extent that objects referred to have the properties referred to. However, in §1.2.1, we saw that metaphysical questions of vagueness can be separated from semantic ones. That means that it should be possible to give a meaning representation with a fuzzy logic that does not assume that there are properties in the world like redness that are themselves vague.

I do not wish to defend a fuzzy view, but we should nonetheless be clear how a fuzzy position would work, given that any worldly vagueness had been idealised away. A traditional view of colour terms might hold that there are many colour properties that some object might have, and still be green. For other colour properties, if an object has them, it would count as greenish/not green. As we have seen, both epistemicist and multiple proposition theorists seem to characterise what is expressed by uses of sentences containing vague terms as relating to multiple properties to capture the one-many relationship between predicates and properties.

A fuzzy account where the proposition expressed receives a fuzzy truth value could be composed of a vague predicate where the predicate would be represented as a fuzzy set of entities. For 'green' these entities would each have some colour property and would be within the fuzzy set to some degree.

Alternatively, a richer account might have the fuzzy set composed of colour properties such that any two objects with the same colour property would be a member of the fuzzy set to the same degree. Either way, no vagueness in the world needs to be incorporated into the view.

To my knowledge, neither of the above options have been explicitly developed as semantic accounts of vagueness and the details of any such accounts might leave a large number of possible treatments of the meanings of vague terms open.

Aside from the complexities arising from his plurivaluationism, Smith's account is semantically fairly straightforward. Properties are fuzzy, so predicates are too. In chapter 2, I will give some reasons to doubt this simple picture of predicate/property correspondence. In chapter 8, I will consider reasons why fuzzy degrees are less desirable than Bayesian degrees.

One fuzzy degree theory that I will consider later will be MacFarlane's fuzzy epistemicism. He takes meaning representations to be probability distributions over fuzzy degrees of truth. His formal model will bear some similarities with the model I will adopt. I will consider it after my view has been detailed.

1.7.2 Bayesian Theories

Bayesian approaches to vagueness have numerous possible entry points into the literature. Proposed Bayesian accounts can be found as daughters in all three subdivisions of the tree in §1.3 (daughters of [11], [10] and [0]). However, one could also imagine the integration of probability calculus into other established theories. For example, a probabilistic contextualism ([1111] or [1110]). Below is a brief description of some Bayesian theories.

Augmented Bayesian Theories

Probability calculus, developed in Kolmogorov (1950), is meant to model reasoning in conditions of uncertainty. In (Lassiter 2011), the Bayesian machinery is been applied to model uncertainties over the extensions of a vague term (propositions) when used in an utterance (Bayesian Epistemic Logic [1000]). These approaches tend to assume that there is are multiple possible interpretations of vague terms. Probability calculus is then used to provide some structure over these possible interpretations such that some are more probable than others. These augmented theories tend to inherit a semantics from an alternative approach. The probabilistic aspect to these approaches provides a modelling of the uncertainty we seem to have over what the extensions of vague terms are.

Edgington's Pure Degree Theory

Edgington's pure degree theory (located at [110]) is the only one not to be an [0] daughter. I take this position to be Edgington's for two reasons: (i) Edgington defends bivalence[45]. That rules out a fuzzy-like interpretation. (ii) Edgington denies that verities are the same as credences (Edgington 1997, §6). That seemingly rules out the uncertainty interpretation of verities. However, in chapter 8, I shall argue that verities can be interpreted doxastically.

Edgington (Edgington 1997) suggests that a number of structural similarities hold between verities and credences. Her insight is that, in virtue of these similarities, the definitions of connectives from probability theory can be used to model vague reasoning as effectively as they can be for modelling reasoning in conditions of uncertainty. Rather than the Łukasiewic-style definitions of connectives (§1.7.1), we have definitions of connectives in terms of verities and conditional verities (which are the analogue of conditional probabilities):

$$\begin{aligned} v(\neg q) &= 1 - v(q) \\ v(q \wedge r) &= v(q) \times v(r|q) \\ v(q \vee r) &= v(q) + v(r) - v(q \wedge r) \end{aligned}$$

The adoption of the above axioms into a logic provides a pure degree theory. However, verities are not degrees of truth in the sense of Fuzzy logic's replacement of two truth values with an infinite number. Edgington provides the following argument for separating degrees of truth of 1 from (the absolute notion of) truth ('true_A' denotes the absolute (Boolean) notion):

> "Suppose 'a is red' is definitely borderline: neither clearly true nor clearly false. It would not be definitely wrong to call it true_A; but it would be definitely wrong to give it a degree of truth of 1. Therefore true_A is not to be identified with a degree of truth of 1." (Edgington 1997, p. 299)

Although a verity of 1 is sufficient for truth, there is no strict mapping from verities to Boolean values (Edgington 1997, p. 299).

[45] This is strongly implied in (Edgington 1997, V5): "The neologism serves to emphasise that I do not see verity as disturbing or displacing the concept of truth"

It seems that statements express a single proposition which is either true or false. However, the same proposition could, as well as having an absolute value, also have a verity. Edgington seems to see vagueness as a phenomenon that forces us to give up on trying to apply absolute values (Edgington 1997, p. 299, V6). In that sense, absolute terms are just not fit for purpose in vague contexts. Edgington clarifies her position in (Edgington 2010) in which she takes vagueness (in the sense of non-clarity) as a primitive notion. The sense in which vague terms do not carve out sharp boundaries is the sense in which they admit of unclear cases. That is Edgington's position which I take as approximating position [110] on the theory tree.

However, we could also make use of verities in other ways. We could treat them either as degrees of truth, or as degrees of uncertainty.[46] It is to these alternatives that I now turn.

Other Pure Degree Theories

Two more possibilities for Bayesian approaches defend non-classical propositions (are daughters of [0]). The first has much in common with fuzzy logic ([01]). As we have seen, fuzzy logic could be adopted to model reasoning with vague terms without determining any metaphysical assumptions about whether properties in the world are fuzzy or not. Using the Bayesian probability calculus as a logic for degrees of truth is an alternative to fuzzy logic.[47]

The second kind of Bayesian pure degree theory (positioned at [00]) differs from the last approach in that it treats propositions as probability-functional, not truth-functional (propositions can be seen as functions into probability values, not truth-values). This form of Bayesian pure degree theory is, perhaps, the most radical to be proposed. Rather than building a semantics based on truth, it holds that the meanings of terms can be specified in terms of uncertainty. It is this kind of view that I will develop and defend over the course of this thesis. The only other directly comparable accounts in the literature that I am aware of are the (van Eijck and Lappin 2012) and (Cooper et al. 2013) manuscripts mentioned earlier.

Such a radical departure from the norm will be in need of substantial motivation. Below I will begin to set out the kind of considerations that will lead to this position.

[46] Edgington provides arguments against both of these alternatives.

[47] There is a question as to whether values for complex propositions (involving connectives) would be genuinely truth functional on this approach. I address this question in §8.2.2.

1.8 Impasse and Methodology

As is often the case in philosophy, new ideas, or sometimes even old ideas presented in a new way, can create a flurry of activity and a lot of literature. In the philosophical vagueness literature, the last hive of activity began in the 90s, where most of the excitement surrounded epistemicist theories. A whole raft of arguments for and against epistemicism were presented. Some writers stalwartly defended prior treatments of vagueness such as supervaluationism and fuzzy semantics. Others who once defended these approaches, became convinced by the epistemicist arguments and started to argue against their former viewpoint.

Now the debate is in something of an impasse. As we have seen, different approaches to vagueness typically start out with the Sorites paradox. The paradox has prompted some to propose far-reaching suggestions about the kinds of things that vague terms can mean, or, at least be used to express. Is it that our words express exact propositions where classical semantics is interpreted literally and some use of a vague term has some specific extension? Perhaps instead we unknowingly express more than one of these propositions?

That there is now an impasse should not be surprising. There were likely to be many solutions to the paradox. The question is, which one is right? The answer may depend on a few factors. For example, if classical logic is modified as part of the solution, the new logic may have unsavoury results. Yet, how classical one wants one's logic to be seems to be a matter of taste, as much as anything else. Some treat logic as a tool, where a different logic can be used for a certain task if it fits the job better. Others see the choice of logic as a commitment to a world view.

Another criteria is how well the theory reflects what words mean. Here, however, the general strategy of starting out with the Sorites comes unstuck. If part of the justification for some view of meaning comes from considering solutions to the Sorites, how well a theory models the Sorites paradox is not going to be a good standard of comparison. Furthermore, since different theorists have come up with radically different answers from the same data (the Sorites), it is likely that some pretty fundamentally different assumptions are being made about things like meaning, truth, and the connection between the two.

Consensus within the vagueness literature is hard to find, but there is consensus that vagueness is pervasive in natural languages. Fairly unanimous is the verdict that vagueness gives rise to Sorites arguments and that, whatever one says about vagueness, it should avoid soritical conclusions. Once philosophers started to turn their attention to natural languages, a study of vagueness was unavoidable. The pervasiveness of vagueness and the threat of Sorites to the consistency of natural language reasoning perhaps explains why vagueness became such a pressing issue starting around that time.

Supervaluationist and fuzzy treatments of vagueness can be seen as a response to the threat of Sorites. In effect, a particular form of argument was used as a justification for the adoption of alternative models of meaning and alternative conceptions of truth. Epistemic treatments of vagueness were a backlash against this approach. For them, Sorites arguments provided no reason to abandon traditional models of meaning and truth. Even for epistemic approaches however, a solution to Sorites has been at the forefront of theorists' minds, not only as a problem to solve, but also as a test of adequacy for the result of any theorising that has taken place.

The place of the paradox at the centre of theories concerning vagueness goes some way towards explaining why there has been no consensus formed, and why there seems to be little hope of a breakthrough of one treatment over another: (P1) Focus on solutions to the Sorites obscures deep disagreements over some fairly substantial semantic issues. For example, over what is expressed by utterances containing vague expressions (one vague proposition, one or many sharp propositions etc.). (P2) But there is no reason to think that there is any bound on the limits of possible solutions to the paradox, nor on the kinds of semantic assumptions one could make.

From the (P1) and (P2) it follows that: (C1) We should not expect those with different solutions to the Sorites to agree on what makes such a solution, or indeed an account of vagueness, successful. (C2) Just because something is a solution to the paradox is no reason to suppose that the semantic assumptions it makes are at all informative with respect to the meanings of natural language vague terms.

The way out of this quagmire that I will adopt is to set up some criteria for an adequate account of vagueness in advance of, and independent of, considerations relating to Sorites arguments. Provided that the criteria are sufficiently broad, no single solution to the paradox will be assumed at the outset. Once an account has been developed, a further test of adequacy will be whether Sorites arguments remain problematic.[48]

[48]The criteria can be linguistic without prejudging how vagueness will be characterised. On any of the major accounts, vagueness arises from words or uses of words. The point of dispute is whether vagueness itself is linguistic.

The two kinds of considerations I will take as central will be (i) language as a tool for communication, and (ii) language learning. Language is for communication, and communication enables us, amongst many other things, to perform tasks and coordinate our behaviour. All this we do in a very complex world with relatively few words. Whatever we say about vague terms, it should be compatible with pressures arising from communication. Before we are able to use language for the purposes of communication, we have to learn language first. Whatever we end up saying about the meaning of vague terms, a further adequacy condition will be that they must be learnable. This is not to say that the extensions of terms, if there are any, will be learnable. Rather that, whatever it is that we learn when we become competent speakers must be something that we plausibly could learn.

In chapter 2, I shall argue that these criteria militate against a truth-conditional view of meaning and thereby a truth-conditional modelling of semantics. If that argument is right, then it suggests that a lot of characterisations of vagueness above have got off on completely the wrong foot. In chapters 3-5, I shall argue that considerations of learning and communication suggest that uncertainty rather than truth should be our central semantic notion. It will be suggested that if that is the case, then the right kind of tools for modelling semantics will be probabilistic. In chapter 6, a semantic model will be proposed and evaluated against some similar approaches in the literature. In chapters 7 and 8, I will return to the question of truth. I will suggest that a probabilistic account of meaning provides a grounding for a notion of degrees of truth and indeterminacy. In chapter 9, I will address the Sorites paradox. It will be used as a test to see if the kind of account of meaning proposed is susceptible to the paradox. Final conclusions will be drawn in chapter 10.

2

Truth, Truth-Conditions, Meaning & Semantics

There is an orthodoxy in the philosophy of language to appeal to truth and truth-conditions, both as a part of laying out a theory of meaning, and as a part of developing a natural language semantics.[1] In semantics, the position is put succinctly by (Lewis 1970): "Semantics with no treatment of truth conditions is not semantics". On the side of theories of meaning, McDowell echoes a common theme: "The thesis should be, not that sense is what a theory of truth is a theory of, but rather that truth is what a theory of sense is a theory of" (McDowell 1976, p. 8).[2] The aim of this chapter will be to undermine this orthodoxy.

In §2.1, I shall examine some of the main theories of truth that have been proposed. I shall argue that all but one of these theories carries a commitment that truth and truth-conditions cannot play an explanatory role in theories of meaning. The exceptions are what I shall call *logical correspondence theories*. These theories are grounded in a Tarskian semantic definition of truth. Different versions of these theories gave rise to two major

[1] Given the different ways that the terms 'meaning' and 'semantics' are used within the field, I do not exclude the possibility that providing a semantics for a language involves, or is even identical with, providing a theory of meaning (for that language).

[2] Other statements of this kind of view are not hard to find. Famous examples include (Frege 1892/1948): "Every declarative sentence concerned with the referents of its words is therefore to be regarded as a proper name, and its referent, if it exists, is either the true or the false." and (Davidson 1967): "to give truth conditions is a way of giving the meaning of a sentence" and "A theory of meaning for a language L shows 'how the meanings of sentences depend upon the meanings of words' if it contains a (recursive) definition of truth-in-L. "

strands of research in the philosophy of language: Davidsonian theories of meaning, and formal semantics in the Montague-Lewis tradition. In §2.2, I will detail the differences between these two approaches to truth and meaning. I will criticise a particular form of Davidsonian theory of meaning with respect to what it entails for semantic learning. Finally, in §2.3, I will claim that the only viable option left, the Montague-Lewis model, makes idealisations that are untenable. The conclusion will be that an alternative representation of the meanings of sentences and expressions needs to be developed.

2.1 Truth and Truth-Conditions in Explanations

Many people have disputed Austin's claim that statements are the (only) bearers of truth. Other candidates that were rejected by Austin, but have been defended by others, include sentences, propositions, and beliefs. Our focus is language, so I will not consider the last of these. That leaves statements (roughly, utterances), sentences, and propositions. Some theories of truth are very minimal and claim that there is very little that can be said about the truth of a statement/sentence/proposition. Others are more substantial in their explanations.

Before going into the details of these theories, it is worth noting that what these theories of truth are theories of (be it sentences, statements or propositions), are all things which have, or perhaps even are meanings.[3] The natural question to ask, therefore, is where the meaning part and the truth part relate to one another in terms of explanations. If it turns out that all the ways we have for spelling out a theory of truth rely in some way on something (the thing for which the theory is defining truth for) which is meaningful, we had better be careful not to rely on truth at the wrong point in our theory of meaning lest the theory becomes circular.

I will consider two kinds of truth theory, varieties of which are the most popular within recent philosophy. The first kind are deflationary theories. Deflationism has its modern roots in Frege (Frege 1918/1956) and Quine (Quine 1970/1986).[4] Quine and Frege argued that the truth predicate is disquotational, roughly, that to say that 'p' is true, is just to say/assert/commit oneself to p. In Tarski, too, we can find a formulation (but not a defence) of deflationism. Tarski (1944) raises and dismisses

[3]If, that is, propositions are taken to be meanings.
[4]Also in (Ramsey 1927) and (Ayer 1935).

an account on which 'truth' is undefined in the metalanguage (such that the fundamental properties of truth are given as a series of axioms).[5] In opposition to deflationism there are a large variety of inflationary theories. Some of these, such as coherence theories, have fallen out of favour and will not be discussed here. By far the most prevalent inflationary theory of one form or another is correspondence theory. The basic idea behind correspondence is that truth consists in a link between words and the world, and the job of a theory of truth is to provide a theory of this link. There are two kinds of correspondence theory: (what I shall call) *speech act correspondence* (Austin 1950/1979) and *logical correspondence* (Tarski 1944), (Davidson 1969), (Field 1972). The former provides an account of truth in terms of linguistic and social conventions. The latter has its roots in Tarski's semantic definition of truth.

The claim I will make is that deflationary theories and particular kinds of inflationary theories are committed to one of the following: (i) Truth is excluded from playing substantial explanatory roles which debars truth from playing an explanatory role in providing a theory of meaning. Knowledge of truth conditions is not equivalent to knowledge of meaning. (ii) A theory of meaning is presupposed in a theory of truth (meaning plays an explanatory role in a theory of truth), and knowledge of truth conditions is not equivalent to knowledge of meaning. If anyone wishes to adopt a theory of truth that has either of these commitments, then they cannot then adopt a theory of meaning that appeals to truth and truth-conditions.

I will consider deflationism as found in (Field 1994) and (Horwich 1990), and the two types of correspondence theory.[6] I shall argue that deflationism is committed to (i) and speech act correspondence to (ii). Logical correspondence will be the odd one out since it isn't clearly affected by concerns over orders of explanation. It will be addressed at the end of this section and the entirety of the next.

2.1.1 Deflationism

Tarski's T-schema: 'X is true iff P', was originally suggested as the goal for a definition of truth: "We shall call a definition of truth 'adequate' if all these equivalences follow from it" (Tarski 1944, §4). Tarski had his own way of

[5]Tarski only mentions this as a possibility for a definition of truth in the metalanguage when the condition of essential richness is not satisfied (Tarski 1944, p. 352). He does not deny that such an account could be useful for some purposes.

[6]I will have something to say about pragmatism in chapter 7.

providing such a definition, but more on this later.

Modern deflationists ((Field 1994), (Horwich 1990)) adopt a suggestion that Tarski rejects (detailed above). They view the T-schema, not as a schema for a series of theorems, but as a schema for a series of axioms. As such, they treat 'true' as an undefined notion in the metalanguage. Substitutions into the schema yield axioms in a theory of truth. Defining truth would be a matter of listing every axiom (an impossible task), hence the utility of the axiom schema. Beyond the statement of these axioms, deflationists deny there is anything more that can be said about truth. Substitution of a sentence into a schema states a truth condition for that sentence.

Deflationist truth is a simple notion about which very little of substance can be said. Deflationists strongly deny that other philosophical notions such as meaning can constitute part of the theory of truth. Although no philosophical notions can play a role in describing truth, truth can play a part in formulating other philosophical principles. For example, Horwich (1990) and Field (1994) suggest that the truth axiom can be combined with principles of first-order logic and deflationary axioms relating to reference and satisfaction to derive principles such as:

$F(a)$ is true *iff* a refers to x and x satisfies F

Since there is nothing more to say or know about truth than the axiom schema (or any substitution into it), all truth can do, as part of some explanation, is take us from one side of the biconditional to the other. It will be useful when combined with other logical or deflationary principles to infer others, but that is it. Truth, on this view cannot play any substantial role in a theory of meaning.

It should not be surprising that deflationists downplay the role of truth conditions in explanations, too. Horwich, at least, allows that some substitution for P into the axiom schema may be called a truth condition for P. However, on Horwich's account, truth conditions are derived from the meanings of the terms in the utterance in question and from a grasp of the meaning of 'true' (Horwich 1998, p. 73). For Horwich, meanings are given as part of a use theory of meaning that does not include a notion of truth. He therefore explicitly denies that knowledge of a truth condition can, even in part, constitute knowledge of the meaning of P because knowledge of meanings play a role in deriving knowledge of truth conditions.

The point is supported by Field. Knowledge of meaning (here, 'understanding'), is not related to knowledge of a truth condition as an identity. Rather, which truth conditions one can know (what one can apply 'true' to) is restricted by what one understands:[7]

> "A person can meaningfully apply 'true' in the pure disquotational sense only to utterances that he has some understanding of; and for such an utterance u, the claim that u is true (true-as-he-understands-it) is cognitively equivalent (for the person) to u itself (as he understands it)." (Field 1994, p. 250)

For both Field and Horwich, truth is excluded from playing an explanatory role in a theory of meaning and knowing a sentence/proposition/statement's truth condition is not equivalent to knowing its meaning. Deflationism is committed to (i).

2.1.2 Correspondence Theory

Inflationists hold that more substantive claims can be made about truth. T-schemas can similarly been seen as goals for a definition of truth, but, involved in that definition, appeals to other, possibly semantic, notions may be made. Correspondence theories are inflationist. They say more about truth than simply stating substitutions into Tarski's schema as axioms of a theory. Correspondence theories start out with the basic intuition that propositions/sentences/statements are true because they/what they say is connected with the world in some way. Early forms of correspondence theories ran into trouble. As good as the slogan 'S is true if it corresponds with/fits the facts' sounds, spelling out what (the) facts are and just what kind of relation correspondence is proved a very difficult task.

The two forms of correspondence theories I will consider here both try to abstract away from facts talk and correspondence talk. The first, speech act correspondence, was advocated by Austin (Austin 1950/1979). It may be surprising to hear Austin's theory described as one which abstracts away from talk of facts and correspondence, since his work on truth is best known as part of an exchange between him and Strawson.[8] In this exchange, much of the discussion centres around a discussion on facts. Nonetheless, Austin's account of truth is made without the use of 'facts' or

[7]This is a somewhat puzzling idea in Field. I might believe u to be false, in which case, we should not expect an utterance of 'u' to be cognitively equivalent to an utterance of 'u is true'.

[8]The full exchange takes place over four papers (Strawson 1949), (Austin 1950/1979), (Strawson 1950) and (Austin 1961/1979).

'correspondence' and he explicitly stated his dislike of the terms, even if he saw more that a pearl of truth in the idiom 'fits the facts'.

The second form of correspondence theory, logical correspondence theory,[9] abstracts away from correspondence talk even more than Austin. In ways to be detailed below, the key notion for logical correspondence theorists is *satisfaction* of a sentence relative to an assignment.

Whereas speech act correspondence will be argued to presuppose a theory of meaning in a theory of truth and thereby be committed to denying that knowledge of truth conditions are equivalent to knowledge of meanings, logical correspondence will not clearly fall into category (i) or (ii).

Speech Act Correspondence

Austin's speech act correspondence appeals to conventions. Sentences conventionally correlate with types of situations/states of affairs. When a sentence is used to perform a speech act, if conditions are right, different conventions correlate that act with some actual state of affairs.[10] The words used to perform a speech act are true if, and only if, the state of affairs correlated with the use of those words is "of a type" with the situations that the sentence is correlated with conventionally (Austin 1950/1979, pp.121-2). What is meant here is simply that truth is determined by looking at whatever part of the world (situation) is in question and determining if it is of a type with the type of situation that the words, meaning what they do, conventionally correlate with. Meanings (of sentences and uses of them), on Austin's account, are provided, at least in part, by conventional correlations. In speaking we can refer to actual situations. If the actual situation we refer to is of the type of situation that our words conventionally correlate with, then we have said something true.

Notice here that nothing has been said about 'being of a type with'. The complexity and subtlety of Austin's view of truth is in this notion. Being of a type is described as being "sufficiently like the standard states of affairs with which" (Austin 1950/1979, p. 122, fn 2). This allows for a great deal of flexibility, and, as we shall see in chapter 7, vagueness.

The details of Austin's account are not vital. Whatever might be put in place of conventional correlation, first one must establish the meanings of

[9]I coin this term so as to avoid Tarski's own description, 'semantic definition of truth'. This is not because I view Tarski's naming of the account as flawed in some way. Rather that theories to be discussed under the label 'logical correspondence' are not clearly Tarskian theories. The connection between Tarski's semantic definition and others' interpretations of Tarski will be discussed at length below.

[10]Of course, this may fail when there is no situation referred to.

sentences. These meanings are characterised as types of situations. Truth is defined as a relation between these situation types and actual situations. Since truth is defined in terms of a relation for which meanings (types of situations) and actual situations are *relata*, truth cannot itself be central to characterising meanings.

Other neo-Austinian views have similar outcomes. The notion of truth in Situation Semantics and Situation Theory[11] was inspired by Austin. In brief, sentence meanings are interpreted as constraints that link utterance types with types of situation. Utterances represent situations as being of some particular type (for example, one in which Mary runs). Utterances are true only if the situation they represent as being of some type is, in fact, of that type.[12] Again, meanings play a critical role in arriving at determining whether an utterance is true. Truth cannot therefore be a central notion in spelling out meanings.[13]

On Austin's view, knowledge of a truth condition of an utterance is not the same as knowing the meaning of an utterance. That is because one can know which situation is being described and what type of situation one's words conventionally correlate with, and yet still not know what it takes for the former to count as being of a type (sufficiently similar) to the latter. There is, however, a one-way implication. If one knows under what conditions an utterance would be true, one must know the meaning of that utterance.

On some modern interpretations of Austin, the complexity of 'being of a type' is not so clearly present. For example, in (Devlin 2006), there is little distinction drawn between the propositional content of an utterance, the information conveyed by the utterance and the circumstances in which the utterance is true. Utterance meanings can be glossed as the proposition that a situation is of some type. The truth condition for an utterance can be glossed as the situation being of the relevant type. Seemingly, knowing one may count as knowing the other. Nonetheless, this simple mapping can be seen as an idealisation, not as a claim of reduction (meaning to truth-conditions). Complexity will need to be reintroduced since we should not expect there always to be a simple answer as to whether some situation

[11]The literature is large. For an original formulation, see (Barwise and Perry 1983). A good summary of situation theory and situation semantics can be found in (Devlin 2006). A modern version of a situation theoretic approach that relates to Austinian truth is Type Theory with Records (TTR), details of which can be found in (Cooper 2005) and (Cooper 2012).

[12]Details can be found in (Devlin 2006, §§8-10).

[13]Many aspects of situation theory are used in the formalism that I adopt in chapter 6. I will leave technical details until then.

is of the relevant type.[14]

It should, furthermore, be stressed that the emphasis in situation theoretic views is not on truth, but on the informational channels in our environments and the informational channels connected with the uses of words. Knowledge of meanings of words (either as types or tokens of utterances) are not meant to be simply reducible to a statement of 'u is true iff P'.

On Austin's original conception, knowing the truth conditions of an utterance and knowing its meaning can come apart. Situation theoretical approaches remain Austinian in spirit. Both are committed to (ii).

Logical Correspondence

Logical correspondence employs the following strategy. Simple formulas can be defined as true relative to an assignment of values to the free variables of the language (a satisfaction of a formula by an assignment). The satisfaction of more complex formulas is defined recursively. Closed sentences (formulas with no free variables) are then true on all assignments or false on all (it doesn't matter which assignment is chosen). Truth of a sentence is defined as truth on all assignments (satisfaction of a sentence by all assignments).

The question then becomes whether meaning is being assumed. This semantic approach to truth applies within a model (it defines truth relative to a model). A model is, minimally, an interpretation function and a domain $\langle D, I \rangle$. Further recursive machinery allows the building of interpretations of complex formulas from simple ones (interpretations of sentences from interpretations of predicates, variables and names). On some understandings of 'meaning', truth relative to a model is truth relative to a specification of meanings for terms (denotations of names, ranges of applications of predicates over a domain).[15] If that is so, then perhaps logical correspondence is in the same boat as speech act correspondence. There is a difference, however. Logical correspondence may well assume some meaning-notions to define truth[16], but whereas speech act correspondence is not a reductive project (sentence/utterance meaning to truth-values), logical correspondence, to some extent, is. On extensional models at least, logical correspondence uses recursive definitions to compose, from more basic denotations, a contents for a sentence (where contents are functions from contexts to truth

[14] I elaborate this idea in chapter 6.

[15] This is what Field means when he describes Tarski as succeeding "in reducing the notion of truth to other semantic notions" (Field 1972, p. 347).

[16] An assessment of this point will be made in detail in 2.2.

values).[17]

Because different notions of meaning are being employed here, and because the details of different logical correspondence accounts carry different commitments to what role these different conceptions of meaning play in a theory of truth, it is not at all clear that logical correspondence theories can be placed in either category (i) or (ii). That is to say, to what extent any meaning notions are being assumed (and in which way they are, if they are) is a thorny issue.

In §2.2 we will look at various ways logical correspondence theories have been fleshed out. These will fall into neither (i) nor (ii), but in not doing so, they will face problems of their own.

2.1.3 Summary

This section has focussed on what might be described as the more metaphysical questions that arise in theorising about truth and its relationship to a theory of meaning. Analyses of links between words and the world for inflationist theories leave little room for an appeal to truth and truth-conditions within a theory of meaning. In a sense, this is not news. One of the concerns of philosophers in the early-mid twentieth century was that terms like 'truth' and 'meaning' were suspiciously metaphysical and ungrounded. One way out of this quandary is to become deflationary about truth and focus instead on conceptions of meaning independent of a truth definition. For my purposes, it does not matter which, if either of these positions is right. All that I have so far sought to establish is that if either of them are right, then an account of meaning cannot rest upon an account of truth.

A change in the prospects philosophers gave to linking the two began with Tarski. Logical correspondence theory was seen as being an exception to the problems of other older truth theories. One interpretation of logical correspondence is that it introduces the possibility of a reduction of truth to more basic formal relations like satisfaction. This would allow truth (and truth conditions) to play a role in stating a broad theory of meaning. Still to be discussed is: what role this is, what sense of 'meaning' is being used, and whether such an account would be viable.

[17] On intensional models, sentence meaning is associated with intensions, where intensions are usually characterised as functions from indices to contents (where contents are functions from contexts to extensions).

2.2 Logical Correspondence Theory

The only place to start in a full discussion of logical correspondence is with Tarski. Tarski's work has been hugely influential in the philosophy of truth and the philosophy of language. The kinds of positions that have been developed from his work are diverse and certainly not consistent. Within the philosophy of language, recursive definitions of meaning and truth have been employed within the Montague-Lewis tradition of formal semantics, and a similar apparatus has been used to different effect within the Davidsonian tradition. Within the philosophy of truth, Tarski has been cited as the inspiration behind both inflationary and deflationary theories of different sorts. It is widely acknowledged to be less than clear what Tarski himself saw as the philosophical connotations of his own work. In some parts of his writing, Tarski makes some fairly grandiose claims, whilst at other times, his goals appear to be more modest. The following two quotes provide this contrast:

> "What terms are we to use in constructing the definition of truth? In the course of these investigations I shall not neglect to clarify this question. In this construction I shall not make use of any semantical concept if I am not able previously to reduce it to other concepts." (Tarski 1956, pp. 152-3)

> "The simplest and the most natural way of obtaining an exact definition of truth is one which involves the use of other semantic notions, e.g., the notion of satisfaction." (Tarski 1944, p. 345)

These two claims are by no means inconsistent, but they differ greatly in ambition. Many people working within formal approaches to semantics may well be inclined to accept versions of the latter without wanting to make any claims about reduction.

I only gave a very precursory gloss on Tarski's account in §2.1.2. Since elements of Tarski's work will crop up again and again, but since there is no consensus over what Tarski himself took these elements to be doing within a broader philosophical context, I will take a few pages to give a (non-original) interpretation of Tarski, and also to specify which parts of his work are taken by whom to be saying what.

2.2.1 Tarski's Theory of Truth

I will explicate Tarski's account of truth using a very simple example of a few sentences of a simple propositional language (roughly as it is presented in (Tarski 1944)).[18] More complex examples will be considered in later sections when we turn to different ways that Tarski has been interpreted. Important to note throughout is that Tarski was highly sceptical regarding the plausibility of applying his account to natural languages. His was a concept of truth for formal languages, and particularly for classical first-order languages. Attempts within philosophy to apply Tarski's ideas to natural language will be discussed at length in subsequent sections.

Tarski distinguished sentences from *sentential functions*. Sentential functions may contain free variables, sentences cannot. For our purposes, suppose that basic sentential functions can contain at most one free variable. An example of a basic sentential function might then be 'x is a cat', whereas a sentence (which is also a basic sentential function) may be 'Tom is a cat'. All sentences are basic sentential functions, but not vice versa.

Since the language we will define truth for, L, contains variables, we need a way of interpreting these variables. The standard method of doing this is to use variable assignments. Variable assignments are arbitrary mappings from variables to objects in a domain.[19] Using variable assignments and the notion of *satisfaction*, Tarski was able to define truth in terms of satisfaction relative to an assignment. Sentential functions like 'x is a cat' are satisfied by objects in the domain. For example, say our domain contains only two objects, a and b, the object a may satisfy the sentential function 'x is a cat' but the object b may not.

After defining satisfaction for all basic sentential functions, satisfaction for non-basic sentential functions can be recursively defined. For example, for basic sentential functions A and B, A and B is satisfied if, and only if, A is satisfied and B is satisfied.

Putting assignments and satisfaction together, we get the notion of

[18] I am aware that Tarski exegesis forms a substantial amount of literature. In this chapter, however, I am interested in detailing how Tarski's work has been interpreted within the philosophy of language and formal semantics. I will therefore focus on Davidson's interpretation and Field's interpretation of Tarski. These two authors represent two distinct philosophical traditions. It is these traditions that I will aim to criticise with respect to their placement of truth a central semantic notion.

[19] This is not how Tarski defines interpretations of variables. He used mappings from sequences of objects to positions in sequences. Variable assignment functions and variable assignment variations are a later, and for me, more familiar way of modelling the same thing. I will speak in terms of variable assignments in all of the following, even if this is not strictly correct with respect to Tarski.

satisfaction relative to an assignment. If there are only two objects in the domain, a (total) assignment will map the variable 'x' onto a or onto b. We know that a satisfies 'x is a cat' and b does not. Therefore for any assignment g, 'x is a cat' is satisfied relative to g if, and only if, the object g assigns to x satisfies 'x is a cat'. The recursive definitions of satisfaction for complex sentential functions ensure we have a definition of satisfaction relative to an assignment for all expressions of the language.

We can now appeal to all (possible) assignments. If L only contains one variable, x, and there are two objects a and b, there are only two possible assignments, one which maps x to a and another which maps x to b. Given that 'x is a cat' is satisfied by a but not b, we can therefore say that it is not the case that 'x is a cat' is satisfied relative to all assignments.

The point at which Tarski's definition of satisfaction becomes interesting is when we consider (i) closed sentences (ii) quantified sentences and (iii) open sentences that are tautologies or contradictions. For example, sentential functions with no variables are either true on all assignments or false on all assignments. Truth conditions for sentences with variables bound by quantifiers can be given relative to all assignment variations. Open sentence tautologies are true on all assignments and contradictions are false on all assignments. And that is the definition Tarski gives for true sentences: A sentence is true if, and only if, it is satisfied relative to all assignments.

Material Adequacy and the T-Schema

Prior to providing his semantic account of truth, Tarski sets up an adequacy condition for any such account. This adequacy condition is based, loosely, on Aristotle:

> "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true." (Tarski 1944, p. 343)

To capture the latter, 'is true', part of this definition Tarski gives his famous generalized definition of truth for a sentence. Where X names a sentence p:

(T) $\qquad$ X *is true if, and only if,* p

An equivalence of the form (T) is one where a substitution is made for p and X. A definition of truth is defined as *materially adequate* if all equivalences of the form (T) follow from it. In this sense, the T-Schema is a goal for a definition of truth. It turns out, with a rather long proof, that the definition Tarski gives in (Tarski 1956) is materially adequate.

2.2.2 Tarski, Field's Tarski, and Davidson

It is not clear whether or not Tarski was a correspondence theorist for natural languages, even though he claims to give a clear and precise formulation of the notion of correspondence which he saw as being imprecise and prone to giving rise to misunderstandings (see, (Tarski 1944, p. 343)).

Some authors, heavily influenced by Tarski, have worn their correspondentist credentials more clearly on their sleeve. This will provide us with two forms of logical correspondence theory to analyse.

In *True to the Facts*, Davidson (1969) defends the semantic theory of truth as a metaphysically respectable correspondence theory. In that paper, Davidson only claims that the Tarskian truth definition can tie words to the world in such a way as to replace talk of correspondence. For the details of how Davidson interprets and adjusts Tarski, we will need to look at some of his other work of the period, principally, *Truth and Meaning* (Davidson 1967) and *Radical Interpretation* (Davidson 1973).

In some of his early papers, Hartry Field defends correspondence theory.[20] Useful for our purposes is that in (Field 1972), he links Tarski's work directly to a model theoretic semantics. Whereas some of Tarski's work seems to take on a semantic eliminativist stance (see §2.2), Field carefully separates this element of Tarski from a less controversial and, as he sees it, more valuable part of his work. Tarski's theory of truth, according to Field, should be seen as a reduction of the notion of truth to other semantic terms such as satisfaction, denotation and predicate satisfaction.[21]

Field's Logical Correspondence Theory

Under Field's interpretation of Tarski (what he calls Tarski*), giving a definition of truth looks very much like giving a pretty standard model theoretic semantics. The language, L for which this semantics will be given has names (c_1, c_2,...), variables (x_1, x_2,...), function symbols (f_1, f_2,...),[22] and one-place predicates (p_1, p_2,...). The semantics for L is given as a model

[20] As we saw, Field would later reject this view for deflationism.

[21] That said, Field reintroduced controversy by then proceeding to claim that Tarski's slimmed down apparatus could be used to give a reductive physicalist model of truth and meaning.

[22] These are of type $\langle e, e\rangle$. An example of an English analogue would be 'the sister of' or 'John's'.

which consists of a domain and an interpretation function. The interpretation function assigns interpretations to all non-logical constants (c_i, f_i and p_i). Variables are interpreted only relative to an assignment g. Assignments are arbitrary pairings of variables with objects in the domain. In Tarski* semantics, semantic terms are defined inductively relative to an assignment.

Before giving Field's Tarski* semantics, we will make a quick aside which will be crucial later. Field (1972, pp. 348-9) suggests that he favours a use theory of meaning, but this is non-essential to the correspondence theory he puts forward. What is essential is that the semantics upon which the truth theory rests is determined outside of the model. The interpretation function describes, say, which name denotes which object. That L is best described by one interpretation function rather than another is determined by something else, say, language use.

The semantics for L can be given inductively. What is below is an adaptation of Field (1972, p. 350), but where more modern notion is used, and the definitions are in terms of a variable assignment function. e is a variable ranging over expressions of L. The subscript X_g indicates that X is relative to assignment g. The interpretation function is $[\![\ ]\!]$. a, b are individuals in the domain:

(A) 1. $[\![x_k]\!]_g = g(x_k)$.
2. $[\![c_k]\!]_g = [\![c_k]\!]$
3. $[\![f_k(e)]\!]_g = a$ if and only if for some b,
(i) $[\![e]\!]_g = b$
and (ii) $[\![f_k]\!]_g$ includes the pair $\langle b, a\rangle$

(B) 1. $[\![p_k(e)]\!]_g$ is true$_g$ if and only if
(i) $[\![e]\!]_g = a$
and (ii) $[\![p_k]\!]_g$ includes the pair $\langle a, 1\rangle$
2. $[\![\neg e]\!]_g$ is true$_g$ if and only if e is not true$_g$.
3. $[\![e_1 \wedge e_2]\!]_g$ is true$_g$ if and only if e_1 true$_g$ and so is e_2.
4. $[\![\forall x_k(e)]\!]_g$ is true$_g$ if and only if for each object a and for each $g[x_k := a]$, $[\![e]\!]_{g[x_k:=a]}$ is true$_g$.

Truth is then defined in terms of assignments and truth on an assignment (true$_g$):

(C) A sentence is true if and only if it is true$_g$ for some (or all) g.

Field's Tarski* semantics, modulo inclusion of such things as a type system and an intensional logic, is pretty much a standard semantics that can be found in any model theoretic approach. If foundational problems can be found with applying this basic model to natural language predicates, we may need to make some far reaching changes. I will claim that there are such foundational problems to be found. These will turn on certain idealisations made about how interpretations are assigned to the basic expressions of the semantics.

Davidson's Logical Correspondence Theory

Davidson's use of elements of Tarski's work (to give a theory of meaning by giving a theory of truth) has been highly influential, especially in British philosophy of language. As already noted, Davidson (1969) sees himself as following in Tarski's footsteps in providing a correspondence theory that doesn't use terms like 'facts' and 'correspondence'. I will concentrate on two main papers of Davidson's. The first, (Davidson 1967), is his seminal paper *Truth and Meaning* in which he first discusses his interpretation of Tarski. The second, (Davidson 1973) will provide some extra detail on Davidson's notion of radical interpretation that is mentioned briefly in his earlier work.

For Davidson, a theory of meaning should, non-trivially, provide the meaning of every sentence in the language. Davidson criticises the Fregean programme of giving, as part of a theory of meaning, a theory of sense. Instead, he collapses Quine's distinction between theories of meaning and theories of reference ((Quine 1953, ch 7)) by suggesting that all one needs as a theory of meaning is a theory of reference.

> "Indeed since a Tarski-type truth definition supplies all we have asked so far of a theory of meaning, it is clear that such a theory falls comfortably within what Quine terms the 'theory of reference' as distinguished from what he terms the 'theory of meaning'." (Davidson 1967, p. 310)

> "a theory of truth patterned after a Tarski type truth definition tells us all we need to know about sense. Counting truth in the domain of reference, as Frege did, the study of sense thus comes down to the study of reference." (Davidson 1984, p. 110)

Davidson's bold move was to suggest that Tarski's schema, originally employed to test the adequacy of a definition of truth, could be utilised to give an adequacy condition for a theory of meaning. Instead of demanding of a theory of meaning that it determines appropriate substitutions into the schema (M)[23]:

(M) s means that p

Davidson suggested replacing the intensional 'means that' with an extensional predicate and connective:

(T) s is T if and only if p

Schema (T) is then used by Davidson to give an adequacy test for a theory of meaning:

> "What we require of a theory of meaning for a language L is that without appeal to any (further) semantical notions it place enough restrictions on the predicate 'is T' to entail all sentences got from the schema T when 's' is replaced by a structural description of a sentence and 'p' by that sentence... it is clear that the sentences to which the predicate 'is T' applies will be just true sentences of L, for the condition we have placed on satisfactory theories of meaning is in essence Tarski's Convention T." (Davidson 1967, pp. 309-10)

It is a tricky issue to what extent this differentiates Davidson from someone like the early Hartry Field. Field's Tarski* semantics delivers us an account of truth for a language in terms of other semantic notions such as denotation and predicate application. The slogan for Field could be 'Doing semantics is to define truth recursively'. Davidson talks in terms of a theory of meaning. His slogan could be 'Giving a recursive definition of truth is to give a theory of meaning'. I shall now argue that there is a way of being Davidsonian whilst adopting a Tarski* semantics. However, I will suggest that Davidson himself does not do so. Finally, in this section, I will argue that Davidson's correspondence is untenable.

Davidson described himself as reversing Tarski's order of explanation. Tarski used meaning to give a theory of truth. Davidson used truth to give a theory of meaning. However, on the face of it, given Tarski*, this reversal seems to be blocked. Davidson wants to use the Tarskian truth definition to give, or at least, to provide constraints on, a theory of meaning. On the

[23] Where 's' is a structural description of a sentence and 'p' is replaced by a sentence that gives the meaning of the sentence structurally described

Tarski* view, giving a recursive definition of truth involves reducing truth to other semantic notions. Truth plays a role in semantics as a notion that has been defined in terms of other semantic notions that have non-semantic explanations (in terms of something like language use). For example, a basic feature of Tarski* semantics is that the interpretations of basic (type e) expressions are defined and inductive definitions are given to enable an interpretation of complex (type e) expressions. Further inductive definitions are given for the truth of sentences (relative to an assignment) built out of an application of an expression to a predicate (to form a basic sentence) and for complex sentences (using connectives, quantifiers etc.). However, truth is not truth *simpliciter*, but truth in a language L. Part of specifying the language for which one is defining truth involves defining interpretations for constants in L. If what this informally amounts to is providing a theory of meaning, then Davidson cannot reverse Tarski's explanation.

However, the above argument equivocates between two conceptions of a theory of meaning. The kind of theory that determines basic denotations and predicate applications might be called a theory of meaning, but it is not the kind of theory that Davidson is trying to give. For him, a theory of meaning should "show how the meanings of sentences depend on the meanings of words" such that a meaning is given for every sentence of the language. It might therefore be argued that even if Tarski* semantics appeals to the meanings of words (the denotations of basic expressions and the ranges of applications of predicates), provided that Davidson ends up by providing the meanings of sentences, he is free to use them. Put another way, employing Tarski* semantics shows just how the meanings of sentences (truth values and truth conditions) are built out of the meanings of words.

There is, then, a Davidsonian position where a theory of sentence meaning is provided by a Tarski* recursive definition of truth but the interpretation function which assigns values to constants is fixed by something else. However, this position does not really reverse anything. Tarski* truth definitions appeal to meanings, but not to sentence meanings, and they anyway show how the meanings of sentences are built from the meanings of their parts. This approach to truth and meaning is the backbone of many approaches to formal semantics. Even though it does not suffer circularity problems (defining meaning in terms of truth whilst defining truth in terms of meaning), it does face other problems. I will articulate and press these problems in §2.3.

The above Tarski* picture might be available to a Davidsonian, but it is not the picture that Davidson himself was defending. Davidson rejects what he calls *the building-block theory of meaning* (roughly what we have seen as Field's Tarski and the project of formal, model-theoretic semantics). The building-block account starts with the interpretations of basic names and predicates and builds interpretations recursively. Davidson's top-down holistic approach removes reference (the interpretations of basic names and predicates etc.) from semantic theory. It treats reference as merely a theoretical construct:

> "A general and preanalytic notion of truth is presupposed by the theory. It is because we have this notion that we can tell what counts as evidence for the truth of a T-sentence. But the same is not required of the concepts of satisfaction and reference. Their role is theoretical, and so we know all there is to know about them when we know how they operate to characterize truth. We don't need a general concept of reference as a condition of spinning an adequate theory.
>
> We don't need the concept of reference; neither do we need reference itself, whatever that may be. For if there is one way of assigning entities to expressions (a way of characterizing 'satisfaction') that yields acceptable results with respect to the truth conditions of sentences, there will be endless other ways that do as well. There is no reason, then, to call any one of these semantical relations 'reference' or 'satisfaction'." (Davidson 1977, p. 256)

However, rejection of the building-block, bottom-up approach places Davidson in a comparable position to a semantic eliminativist reading of Tarski which appeals to a theory of sentence meaning by way of a notion of adequate translation (See Field (1972) for details). If this is so, then Davidson must give a story about when one sentence is an adequate translation of another.

The challenge[24] is easily put. The adequacy condition for a theory of meaning (Davidson's sense) is that it provide substitutions for 's' and 'p' in 's is true if, and only if, p' that result in a true biconditional. Tarski's semantics does this by setting out recursive definitions that make use of sentence equivalence when the object language is contained within the

[24]There are other discussions of this challenge. See, for example, (McDowell 1976).

meta-language. When the object language is not so contained, it makes use of adequate translation.

In (Tarski 1956), this appeal to sentence meaning is explicit:

> "for all a, a satisfies the sentential function x if and only if p...
>
> When ... the corresponding explanation of the expression 'a satisfies the given sentential function' is to be formulated wholly in terms of the metalanguage, then in the above scheme we insert for 'p' not the sentential function itself, but the expression of the metalanguage HAVING THE SAME MEANING, and for 'x' we substitute an individual name of this function which likewise belongs to the metalanguage." (Tarski 1956, p. 190) (my emphasis)

Clearly a formulation of truth that relies on generating sentences of the metalanguage that mean the same as sentences of the object language cannot then be used as a theory of meaning, the adequacy condition of which is to do just that.

Davidson was entirely aware of this pitfall (see (Davidson 1967) and (Davidson 1973)). The way he sets up the problem in his 1967 paper is by asking how a theory of meaning could be constrained in such a way as to generate instances of the T-schema such as, "'Snow is white' is true if and only if snow is white.", but exclude instances such as, "'Snow is white' is true if and only if grass is green". Davidson's response comes in two parts. The first is a little piece of methodological optimism, the second sparked a major line of Davidson's research in the following decade or so.

The optimistic line stems from Davidson's holism. The thought is that sentences such as "'Snow is white' if and only if grass is green" may not be eliminable on their own, but that the adequacy condition on a theory of meaning is that it generate true instances of T-schemas for all sentences in the language, and that there must be systematicity in how these are generated. These generality and systematicity constraints should, claims Davidson, deal with the unacceptable examples. He does not give any more details in his 1967 paper, but it is not clear that this optimism is well founded, or even that Davidson himself was satisfied by this response. In fact it is not unreasonable, given the connectedness of their work, to see Davidson following the exact same steps here as Quine (1960). In chapter two of *Word and Object*, we see Quine consider a similar problem:

> "The infinite totality of sentences of any given speaker's language can be so permuted, or mapped onto itself, that (a) the totality of

> the speaker's dispositions to verbal behaviour remains invariant, and yet (b) the mapping is no mere correlation of sentences with equivalent sentences, in any plausible sense of equivalence however loose." (Quine 1960, p. 27)

This consideration leads Quine into his discussion of radical translation. Davidson turned his thoughts to *radical interpretation*.[25] Tarski's appeal to translation was unavailable to Davidson because it involved, by his own lights, a theory of meaning. His development of radical interpretation was an attempt to show how one might be able to interpret another's language without any prior knowledge of the meanings of their terms or of their mental states (belief and desires etc.). If such an account could be successfully given, then Davidson would have a way of restricting instances of T-schemas in such a way as to ensure that only true instances would be generated by a theory of meaning.

It is in this respect that Davidson can be seen to be trying to reverse Tarski's order of explanation. Provided that we can recognise which sentences are taken to be true (by a speaker), then from this we can build a theory of meaning as a theory of radical interpretation.

The theory of radical interpretation that Davidson espouses is unashamedly holistic. In fact, it is only by being holistic that Davidson manages to reverse Tarski's order of explanation at all. He proposes that we can take, as evidence, all sentences held true by a speaker and optimally fit all T-sentences to that evidence.

> "The present idea is that what Tarski assumed outright for each T-sentence [that it satisfies a translation criterion] can be indirectly elicited by a holistic constraint. If that constraint is adequate, each T-sentence will in fact yield interpretations" (Davidson 1973, p. 326)

A full critique of the Davidsonian project is outside of the remit of this chapter. Instead, I shall assess it in terms of the stated aims of this thesis. Vagueness is ubiquitous in natural language, and it should therefore be central to any adequate analysis of meaning. The deadlock found within the vagueness literature was analysed as arising from fundamentally different conceptions of word and sentence meanings and of the role of logic and truth in this literature. The working hypothesis of the thesis is that

[25] See, for example, (Davidson 1967), (Davidson 1973) and (Davidson 1974).

answers to these questions should be found by taking semantic learning and communication as starting points. For example, whatever it is that words and sentences mean, these must be things capable of being learnt, and they must also be consistent with playing a part in a wide variety of communicative goals. It is with respect to the possibilities for learning that Davidson's account is problematic.

It is unclear what kind of account of learning Davidson is offering. He openly distances himself from anything concerned with "the actual history of language acquisition" (Davidson 1973, p. 313). However, it is not clear from this remark that he actually does mean to distance himself from any discussion of language acquisition at all.

The following is a reconstruction of an account of learning taken from Davidson (1973). A learner must develop a Tarskian style theory of truth for the language they are learning. However, unlike a Tarski* truth definition, they will not be able to do this by associating basic expressions with interpretations and building more complex interpretations from these. Instead, the notion of truth must be taken by the learner as basic. Somehow this will enable the learner to have a list of T-sentences, one for each sentence they have been exposed to, plus the ability to recursively generate more.

Davidson provides a three-stage outline for this process. (1) We must try to fit the whole of first-order logic onto the phrases of the new language "in one fell swoop" (Davidson 1973, p. 323), using the sentences taken true by all speakers all the time as evidence. This process is supposed to allow learners to determine the logical types of words and phrases within the language. (2) Then, one must identify all of the statements sometimes held true and sometimes held false (such as those which contain demonstratives) and tally these with those features of the contexts of their use (time of utterance, speaker etc.) that determine whether they are held true or false. (3) Finally, one should use the restrictions imposed on predicates by the first two steps to determine interpretations for individual predicates and terms as well as for those statements for which there is no uniform agreement over their truth value.[26]

This account has two strange features. For one, it is entirely opaque, under radical interpretation, what language learners will be using as their

[26]This kind of idea was already present before *Truth and Meaning* and (Davidson 1973) in (Davidson 1965/1984). Davidson's early work is mostly negative and criticises various other theories (Quine's Carnap's, Church's) on matters such as attitudinal verbs and quotation as being unlearnable. Davidson is light on a positive view there, but we get a hint when he says "Guided by an adequate theory, we see how the actions and dispositions of speakers induce on the sentences of the language a semantic structure" (Davidson 1965/1984).

metalanguage in this process. However, learners are being assumed to have names for object language sentences and translations/interpretations of each name. Second, notice that learning can only take place on this view by learners tracking truth-value judgements of ALL sentences and statements, along with observations about the conditions of their use, and then using these judgements to develop a truth theory (and thereby a theory of meaning) that is optimised to fit all of the data. We must, on Davidson's view, be able to identify the true sentences of the language and reason everything back from there.

This is not a knock-down argument, but it is very hard to see how this kind of holism is at all compatible with a plausible theory of semantic learning. For a start, hearers are expected to track and remember vast lists of sentences, the conditions for their use and various speakers' truth judgements about them all without any interpretation of those sentences. If this isn't implausible enough, the account is also committed to view that learners first pick out full blown logical structures of a language and learn patterns of inference over these structures, all without any interpretation of individual terms or predicates.

One of the attractive features of Tarski* semantics is that complex interpretations are built out of basic ones. By taking truth as primitive, Davidson loses this feature, ends up in a thoroughly holistic picture, and denies himself a plausible account of how one might learn/acquire that language.

However, the alternative faces challenges, too. As Davidson points out (Davidson 1977), it is a non-trivial matter to account for how a learner may acquire interpretations of, for example, basic predicates or classifiers. In the remainder of this chapter, I shall suggest that if the interpretations of basic predicates are understood as functions into truth values, this problem is severe indeed. In chapters 3-6, I will characterise an alternative picture of the meanings of predicates that takes being learnable as central.

In the next section, we will consider the viability of the role of truth and truth conditions in approaches similar to Tarski* semantics. Such systems do not face holism problems, but I shall argue that they cannot accommodate *semantic variety*. The notion of semantic variety (also *occasion sensitivity*), comes from the work of Charles Travis (see especially (Travis 2008c), (Travis 2008b) and (Travis 2008d)). Travis argues that simple predicative sentences are compatible with a potentially unbounded number of truth conditions. This result of Travis' will be analysed and used to make the claim that truth conditional semantics not only faces a learning challenge of its own, but also faces a severe challenge with respect to communication.

2.3 Truth and Truth-Conditions in Formal Semantics

To adopt something akin to Field's Tarski* approach is essentially to adopt a traditional approach to formal semantics in the Montague-Lewis tradition. This alternative route does not face challenges due to holism. However, in this section I will consider an alternative problem: that some of the idealisations it makes are not plausible.

The goal of formal semantics for natural languages is fairly modest. This is hardly surprising, given the complexity of the subject matter. The basic idea within the Tarski*/Montague/Lewis tradition is that NL expressions can be grouped into logical types. These types will then constrain what kinds of expressions can combine together to form meaningful expressions and what type these meaningful expressions will be. For example, if proper names are interpreted as being entities (individuals) and verb phrases are interpreted as being functional types (represented as, say, a set of entities) then combining the two will return an expression of a sentence/propositional type.

There are two kinds of decisions that semanticists need to make. First, they must make well motivated decisions over what kinds of types should be assigned to what kinds of expressions such that the decisions they make do not produce anomalous predictions for the interpretations of expressions for whatever language(s) is/are being analysed. A huge amount of work has gone into providing analyses of language to attend to these kinds of questions. In all that is to follow, I do not intend to dispute any of the work done in answering them.

The second kind of decision that semanticists must make is how to interpret expressions of different types. There is a significant amount of variation in semanticists' approaches to this decision. On a basic extensional model, a verb phrase like 'runs' might be interpreted, relative to some time and place, as a set of entities of things that (are) run(ning). On an intensional model such as Montague's, ignoring various details which relate to tense and number, 'walks' is interpreted as a function from worlds to a function from entities to truth-values. So, on an extensional picture, 'John walks' is interpreted as a truth value, whereas on an intensional picture, 'John walks' is interpreted as a function from worlds to truth values, which, relative to some world, is still wholly extensional in spirit. On dynamic logic models, sentences such as 'John walks' are interpreted as updates on models, for example, as a function that eliminates worlds in which John is not walking from the common ground of interlocutors.

In this section, I shall present an argument for why traditional answers to the interpretations of semantic types get off on the wrong foot. Since a major topic in this thesis is vagueness, I will concentrate on predicate terms. These come in different syntactic categories, such as verb phrases ('runs', 'makes a great quiche') and adjective phrases ('is tall', 'is red', 'is bald').

Formal semantics, of the kind detailed above, can be viewed as making a series of idealizations about language. For example, if, in a model, a predicate is interpreted as a classical set of objects, then we needn't assume that the semanticist is making any metaphysical commitments about the meaning of that predicate. It may or may not be that the meanings of predicates are best represented this way, but it certainly seems to be the case that by making the idealization that they do, we get a rather neat model and a simple way to describe the application of predicates to arguments. However, we must keep in mind that this idealization needn't be more than just that. As we saw in §1.4.1, a failure to adequately mark the distinction between a model and what is modelled is one route into epistemicism via the basic argument. If one assumes, from the start, that classical semantics accurately captures natural language semantics, one is bound to conclude that all predicates, vague or otherwise, have sharp boundaries. But, it is far from clear that we are entitled to start with that assumption. Under the idealization that meanings can be represented as functions into Boolean values, the interpretations of sentences formed by applying an argument to a predicate will be a truth value. Therefore, if there are reasons to think that interpretations of sentences should not be truth values or functions from worlds to truth values, then that will be a reason to think that the interpretation of predicates should not be a function from entities to truth values (or its intensional equivalent).

The argument I will use to undermine this idealization will be tied closely to work done by Charles Travis ((Travis 1978), (Travis 2008d)). This argument will suggest that truth-conditions and truth-values, even relative to some world, are not good candidates to be the meanings and extensions of sentences. So functions into truth values should not be how we interpret predicates.

2.3.1 Occasion-Sensitivity and the Role of Truth

First some familiar vocabulary. Sentences, or at least declarative sentences, on a traditional picture, are used to express propositions (or thoughts). 'Proposition', in this sense, is a technical term, with a technical definition. On a traditional semantics, the proposition expressed by a sentence (or, at least, the proposition that is the interpretation of a sentence) can be individuated in terms of truth-conditions. On an intensional model, since propositions are functions from worlds to truth values, propositions can be individuated as sets of worlds in which they are true. On an extensional model, propositions are harder to individuate (leading to familiar Frege problems).[27] What we do get on extensional models is a T-sentence stating the truth conditions of the sentence. On a Tarski* semantics, T-sentences are implied by the meaning of a sentence as constructed compositionally from the meanings of its parts (so is not itself a statement of their meaning).

We can give an informal characterisation of the meaning of a sentence which will be compatible with both an intensional and an extensional model, and it will be enough to get Travis' argument going. For every sentence (or use of a sentence), a set of facts about the world of utterance can be stated, relative to which the sentence is either true or false. Put another way, provided we are given enough facts about the world of utterance (such that we can determine to some degree of adequacy, which possible world we are in), we will be able to say, relative to our semantic model, whether a sentence uttered in that world (at some time etc.) is true or false.

Travis' argument is aimed at denying this claim. If he is right, then the idealization within traditional approaches to semantics, that the meanings of sentences (uttered on some occasion) is representable as a proposition/a truth-condition cannot be plausible.

To get his argument going Travis uses, what is now, a well known example:

> "A story. Pia's Japanese maple is full of russet leaves. Believing that green is the colour of leaves, she paints them. Returning, she reports, 'That's better. The leaves are green now.' She speaks truth. A botanist friend then phones, seeking green leaves for a study of green leaf chemistry. 'The leaves (on my tree) are green' Pia says. 'You can have those.' But now Pia speaks falsehood." (Travis 2008d, p. 111)

[27] In fact, intensional models do not solve these problems either, certainly not for logical truths, and not for identities between names either if names are taken to rigidly designate.

There is nothing special about this story. In principle, this kind of change of truth-value could be generated for any simple sentence.[28] Examples like this are supposed to show that, whatever one thinks that 'The leaves are green' means, the meaning of those words is compatible with 'The leaves are green' having different truth-values when describing the same object.[29]

On the whole, some aspects of the data such examples provide is not disputed. All can agree, roughly, that it would be reasonable to call Pia's first utterance 'true' and her second utterance, 'false'. Agreement ends here. For example, it is disputed whether her words were literally true (false) on both or either occasion (see Cappelen and Lepore (2005)), and also whether or not such examples can be accommodated within a theory of meaning. For example, one might hold that one utterance expressed the literal meaning, whereas the other expressed an implicated meaning. For now, I will not assume anything one way or the other, merely that Travis cases demand some form of explanation. The rest of this section will evaluate some of these explanations.

There will be two main approaches I will consider. The first is roughly Gricean in spirit. This response says that both utterances have the same literal meaning, but one also carries an implicature which is false. The second kind of approach does not assume that words have one literal meaning. This approach uses either ambiguity or indices to account for a difference in either the proposition expressed by each utterance, or the truth value that the utterances receive. The first kind of response allocates the solution to pragmatics. The second kind of response details a semantic solution.

2.3.2 Literal Meaning and Pragmatic Implicatures

Grice's view was that words have literal meanings.[30] However, he also thought that a large amount of communication rests both on the keeping and breaking of communicative norms (Grice 1989a), and on speakers'

[28]There may be some exceptions when mathematical concepts are introduced. For example, 'Five is a prime number' may not be able to undergo this sort of change.

[29]A common feature of Travis cases is that the object in question undergoes no changes between utterances.

[30]From the introduction to *Logic and Conversation* (Grice 1989a), Grice himself is ambivalent as to whether we should think of literal/conventional meaning in formal terms. I therefore will not, strictly, be arguing against Grice. At issue here is whether one could retain a formalist conception of context-independent literal meaning modelled with a classical proposition, and then use something like Gricean pragmatics to explain away Travis' data.

intentions (for hearers to recognise their intention to communicate that P etc. i.e. *speakers' meaning* (Grice 1957)). The pragmatic response to Travis need not be committed to the details of Grice's account, however. All that is needed is some way of separating literal meaning from non-literal meaning, and the claim that only literal meaning is expressed/is in force for one utterance, but that non-literal meaning matters in the other. I will give a version of this response that will try to remain fairly neutral on the details of how non-literal meanings (implicatures) arise. That view will then be criticised, not on the details of how implicatures are generated, but on the basis of assuming that classical propositions can represent literal meanings.

In the story about Pia, a defender of literal meanings might say the following. Before she paints them, Pia's leaves are (literally) red. When she paints them with green paint, she makes them (literally) green. After all, one way to literally change the colour of something is to paint it. After painting, the leaves are no longer (literally) red. That is why her first utterance is true. This also entails that her second utterance is literally true. However, Pia has done something wrong by saying they are green in the second instance, and we can give a pragmatic story about what she has done wrong. For example, we might think that when her botanist friend asks for green leaves, something other than the literal meaning is implicated. For example, in virtue of asking for leaves for green leaf science, the friend may be implying that she wants leaves that can photosynthesise. When Pia responds that her leaves are green, either she is not being cooperative (taking on the implicated meaning), or else, she is expressing the implicated meaning and saying something false (although not literally false).

If we want to add more theory, we can. The response has a similar structure to one of Grice's examples. In the example, A says "I am out of petrol" and B replies, "There is a garage around the corner". About this example, Grice says:

> "B would be infringing the maxim 'Be Relevant' unless he thinks, or thinks it possible, that the garage is open, and has petrol to sell; so he implicates that the garage is, or at least may be open etc." (Grice 1989a, p. 32)

Similarly, Pia would be infringing the relevance maxim unless she thought, or thought it possible, that her leaves would be of use in some green leaf science.

I do not doubt that Grice's maxims capture something about communication (I will discuss Grice and Gricean pragmatics further in chapter 3 and chapter 7). The problem with the story is that we need to say what literal meaning is. Why should Pia's leaves be literally green when she has painted them? If literal meaning is meant to reflect what a word encodes context-independently, we would not want *the colour that the object is painted is green* to be what is context-independently encoded by 'green'. It makes sense for someone who recognises a Japanese maple by the shape of its leaves to say "Pia's leaves aren't green. They are only painted green". If being painted green is part of the literal meaning of 'green', then that sentence should be literally false. Then again, we cannot go the other way. The colour of paint sometimes does count. If there is such a thing as literal truth, is it not (literally) true that cars come in different colours (as opposed to the uniform metallic grey that they all are under the paint)?

Alternatively, maybe paint colour is not what features as the literal meaning of 'green'. Yet, even *surface colour*, or *colour of appearance* will face similar problems to *colour of paint*. If it is not any of these things that are included in literal meaning, then either (i) we cannot explain why Pia's leaves can be literally red and then literally green, or (ii) what counts as being encoded in literal meaning can vary from case to case.

If (i), then the literal meaning plus pragmatics response is a non-starter since it cannot justify why either or neither of Pia's utterances is literally true (false).

If (ii), then we lose the firm foundations of literal meaning that were meant to ground the response. If the colour something is painted sometimes counts and sometimes does not count for determining the colour that it literally is, then, pragmatic story aside, it looks as though we'll be able to construct a story where, using the same sentence to speak about the same object, it will be true that the object is literally green on one utterance, and false that it is literally green on another. This amounts to requiring, on top of any pragmatic story, an account of how even literal meaning can change with context. If literal meaning changes with context, it is not anymore literal (context-independent) meaning. This reduces the pragmatic response to one of the contextualist responses that we will consider next.

2.3.3 Context Shifts

If it is right that the meaning of 'The leaves are green', spoken of the same leaves on different occasions, is compatible with more than one truth value, then there is a tension with the traditional semantic model. On the traditional model, if an object (some collection of objects) satisfies a predicate (possibly relative to some world of evaluation), then the very same object (collection) cannot cease to satisfy that same predicate provided that it has undergone no change.

This is a tension, not a contradiction. Different semantic models have various devices to explain this kind of phenomenon. These fall into three main categories which are listed below. To outline these I adopt distinctions made in Kaplan (1989) for the treatment of indexicals. Any sentence expresses a proposition, but this may vary depending on some difference in the context of utterance.[31] Once content is fixed, it can be evaluated for truth in a context of evaluation.

1. The proposition expressed differs on each occasion. This is because:
   (a) The sentence is parameterized or contains a free variable and the context of utterance has changed.
   (b) The sentence is semantically ambiguous, and the context of utterance has changed.
2. The same proposition is expressed by each utterance of the sentence, but the context of evaluation has changed.

Responses to Travis have focussed on (1a). For example Predelli (2005) introduces parameters into his semantics. Although I am not aware of any response to Travis along line (1b), it is possible that a supervaluationist might take this route. There may be many possible propositions expressed by 'The leaves are green'. Supervaluationism can appeal to a change in the context of utterance altering the range of possible propositions being expressed. A response to Travis along line (2) can be found in (MacFarlane 2009a).[32] Anyone averse to the changes in semantics made in responses (1a) and (1b) could keep a conservative approach to semantics and try to account for Travis' phenomena in that way.

In his preemptive reaction to responses to the Pia case, Travis has direct arguments against (1a) and (1b). He seems to take these to be conclusive. I agree with these arguments, but I shall claim that they only succeed in making the force of Travis cases rest on there being no successful defence in the form of (2). I will summarise Travis's criticisms of (1a) and (1b), and I will argue that these criticisms force (1a) and (1b) to, ultimately, rest on a

[31] I am allowing for a broader notion of context of utterance than Kaplan. My notion includes whatever it is that determines semantic disambiguation.

[32] MacFarlane does not defend this position (which he calls 'non-indexical contextualism'), but merely raises it as a possible theory that one could adopt.

defence of (2). In brief, the argument will be that, even relative to a single context of utterance, truth conditions are not fixed. Finally, I shall raise problems for any account of the form of (2) based on its implications for learning and communication.

Parameters and Ambiguities

Travis' response to (1a) comes in two parts. The first argues from a dis-analogy with other indexicals such as 'I' in English. With such indexicals, describing a change in meaning of 'I' from occasion to occasion seems to require making a reference to who is speaking at the time. The fact that 'I' seems, most, if not all, of the time to refer to the person speaking is a reason to think that part of the semantics of 'I' could be captured as a speaker parameter. The value of the parameter then contributes to the meaning of 'I' when uttered.[33] Expressions such as 'is green' do not seem to work the same way. Parameters for 'is green', if they were to work analogously to 'I' would have to be things like *the things that are painted green* and *the things that are naturally green such that they would be suitable for green leaf science*. Neither of these parameters naturally suggests itself as part of the explanation for the meaning of 'is green', at least not in the way that the speaker does for explaining the meaning of 'I'.

Second, Travis describes a way of generalizing indexicality. This can again be understood in the context of Kaplan. On Kaplan's semantics, utterances express contents. For the indexical elements of the sentences uttered, the character of the indexical makes use of some parameter to give that element content. Thus, if Bertrand Russell said, 'I am a lord', the content of his utterance (ignoring complexities relating to tense and the 'am a lord' part) could be given as *Bertrand Russell is a lord*. Call the sentence 'Bertrand Russell is a lord' the generalized sentence of his utterance 'I am a lord'. Unlike the sentence 'I am a lord' (again ignoring tense etc.), the generalized sentence is assumed to have a particular content which will be fixed independently of the context of utterance. Also assumed is that the content of the generalized sentence is the same as the content of the sentence 'I am a lord' as spoken by Russell. Travis' point, inspired by Wittgenstein, about 'The leaves are green' is that for any generalized sentence such as 'The leaves are painted green', exactly the same kind of example could be generated. In this new example, all we would need to

[33]Put another way, 'I' has a character that could be informally characterised as, *the speaker of the utterance in the context*.

do is tell a story where, for two utterances of, say, 'The leaves are painted green' (that express the proposition *the leaves are painted green*), Pia says something true for one and something false for the other.[34]

To put the point another way, a Travis case can be constructed for any content fixed by the context of utterance. If this is the case, then unless this Travis case can be explained by a change in the context of evaluation for that content, no truth-value change should be possible. The result of Travis' criticism of (1a) is therefore to require it to appeal to (2) in the face of further Travis cases.

Travis' criticism of the ambiguity response, (1b) does not take supervaluationism into account. Instead it targets a simple ambiguity theory. Again, it rests on disanalogy:

> "If words are ambiguous in English, there must be a way of saying just what these ambiguities are; so a fact as to how many ways ambiguous they are. The pair of speakings we considered differed in that each invoked a different understanding of what it would be for leaves to be green. There is no reason to think that there is any limit to possible understandings of THAT, each of which might be invoked by some words which spoke on that topic." (Travis 2008d, pp. 112-3)

For other semantically ambiguous expressions, we have strong motivations for positing an ambiguity in an expression (see §1.6). For example, the expression 'is green' is ambiguous. Cars come in many colours, some of them are green. Equally though, few cars are green due to the amount of gasses they emit which are destructive to the ozone layer. The question, "When is a green car not a green car?" has a perfectly sensible answer – "When it's a gas guzzler". We cannot, it seems, play the same game with Pia's leaves: "When is a green leaf not a green leaf?" – "When it's painted green".

Further to disanology, Travis' second argument against (1a) applies to a simple ambiguity account, too. If 'is green' is ambiguous, but any use of it still expresses a single proposition, then another Travis case can be constructed for that proposition expressed. Relative to a single context of utterance, 'green' might be disambiguated to 'painted green', however:

> "Suppose [the leaves] are painted, but in a pointillist style: from a decent distance, they look green, but up close, they look mottled.

[34] See (Travis 2008d, pp. 112-3) for suggestions of how to construct such a story.

> Is that a way of painting leaves green? It might sometimes, but only sometimes, so count. So there would be two distinct things to be said in the presumed *paint counts* sense of 'is green'. And so on." (Travis 2008d, p. 112) (my typeface conventions)

The 'and so on' indicates that a position (1b) cannot plausibly go on claiming ambiguity indefinitely. This will not create an insurmountable problem provided that at some point, changes in truth values can be accounted for as changes in the world of evaluation for the disambiguated proposition. So, in the long run, a (1b) account will need to appeal to a change in truth value in the manner of a (1b) account. Again, a successful (1b) account rests on a successful (2) account.

Perhaps an ambiguity account would not wish to be committed to there being a single proposition expressed in a single context of utterance. That would be the start of either a supervaluationist or a plurivaluationist interpretation. Supervaluationist semantics bites the incredulity bullet over there being (possibly vast) numbers of possible interpretations for sentences. For any context of utterance, a subset of these is expressed. The utterance is (super)true if and only if it is true on all of the propositions (sharpenings) expressed relative to some world of evaluation.

As an aside, independently of how well we take supervaluationism to model vagueness[35], it seems unlikely that a supervaluationist would allow that either of Pia's utterances are supertrue. If so, this would mean that supervaluationism would wrongly predict that Pia's utterances would lack a truth value. For this not to be the case, all sharpenings of 'is green' on which it is false that russet (but painted) leaves are green would have to be excluded by the context of utterance of Pia's first utterance, and all sharpenings of 'is green' on which it is true that russet but painted leaves are suitable for green leaf science would have to be excluded by the context of utterance of Pia's second utterance. Even if we allow the supervaluationist the possibility that expressed sharpenings are restricted in the right way, Travis' second argument against (1a) can be brought to bear.

Relative to a single context of utterance, for a (1a) view, a sentence expresses some particular proposition. On a supervaluationist account some set of propositions is expressed. Each proposition can be regarded as a set of possible worlds. For an utterance to be supertrue, every expressed sharpening must be true. Hence the expressed set of propositions can be regarded as the intersection of the sets of possible worlds that are

[35] See §1.6 for discussion.

equivalent to each proposition. The intersection of a set of worlds is a set of worlds, and therefore a proposition. So relative to a context of utterance, supervaluationist treatments still treat the utterance of 'The leaves are green' as expressing a single proposition. If a Travis case can be constructed that suggests that this proposition can change in truth value, then this can only be accounted for as a change in the context of evaluation. Again, this means that a successful type (2) response must be given to account for the data.

As another aside, other non-classical semantic treatments such as fuzzy logic and nihilism would not help here either. As Travis notes, (Travis 2008d, p 113), Pia's utterances are being assumed to be true (false), not indeterminate. As such, a single proposition, or a completely true/false fuzzy proposition, expressed by each utterance, will be targeted by Travis' argument as effectively as on a more traditional semantics.

Changes in the Context of Assessment

The question of whether a response of kind (2) can account for Travis cases is really a question as to whether the meanings of utterances can be adequately modelled as functions from circumstances of evaluations into truth-values.[36]

The position we are now in is: Relative to some context of utterance, the use of some words, such as 'The leaves are green' has some meaning. On the traditional view this is to say that it expresses a proposition. For an extensional, traditional view this is a disastrous result. On that view, propositions (the meanings of utterances) are functions from contexts to truth values, and so it cannot be that the same proposition can have different truth values relative to a single context of utterance. Travis cases seem to show that they can, and so whatever the meaning of Pia's utterances of 'The leaves are green' is, it can't be an extensional contents.

On an intensional view, there is still room for manoeuvre. If Pia's utterances have, as their meaning, a proposition which is a function from worlds to (contexts to) truth-values, then it is possible that by keeping this content fixed, sometimes 'The leaves are green' is true, and sometimes false, even relative to a single context of utterance. The difference will have to be explained by a difference in the circumstances of evaluation of this same proposition between cases. If this cannot be explained by there being a difference in the circumstances of evaluation of the proposition, then we have reason to doubt that the meaning of 'The leaves are green' can

[36] One such account in the literature is (MacFarlane 2009a).

be accounted for as a single function from such contexts to truth values. If we have reason to doubt that, then we have reason to doubt that the interpretations of predicates like 'is green' can be modelled as anything like a function from worlds to a function from entities to truth values. If that is so, then an alternative analysis for how we represent the meanings of predicates is called for.

Now we can abstract away from the Pia case and examine what learning the meaning of such utterances consists in and how such entities might be used in communication.

Learning

To be competent with a sentence, on the view under consideration, is to be able to say, given sufficient facts (a world of evaluation), whether the proposition that sentence expresses would be true or false. If the sentence is a simple predicate-object construction, this amounts to having to learn, for some predicate, whether it applies to some object in that world of evaluation. Even with vagueness and borderline cases bracketed, this creates a learning problem. If we restrict ourselves to a single predicate, the number of worlds of evaluation that a learner will be exposed to is finite and probably very limited. The learning task would involve coming to know whether the current proposition is true of the world we are in.

If this knowledge is simply an independent fact, unrelated to other facts in the world, then there could be no way to learn it. So, instead, we might have to learn some rules of use, that, if certain other conditions obtain, then green applies/doesn't apply. The prospects for this strategy seem poor, however, when we look back to a specific case. Pia's first utterance is true. The leaves are green even though they are russet leaves painted green. Hence, that they are painted cannot be taken to contradict the claim that they are green. Perhaps knowing that painted green leaves are not suitable for green leaf chemistry is enough to differentiate the two utterances. If so, then as a part of learning how to use 'green' one would learn a declarative rule of inference: `Painted green + suitable for green leaf chemistry → Not-Green`. If this is the story, however, then Travis cases can be brought to apply to that declarative rule that has been learnt. Suppose that Pia has painted her leaves with a chlorophyll based paint and her friend's green leaf chemistry aims at extracting chlorophyll. Now, Pia's utterance could count as true, but the rule just given must be updated or replaced by some new rule or set of rules and so on *ad infinitum*.

This *ad infinitum* marks the impossibility of learning the rules for applying 'green'.

The trouble here stems directly from the idealization that meanings are functions from sets of conditions (worlds) to truth values. By definition, this creates a learning task where the learner must come to know substitutions into the schemata: If X, then apply 'Green'. And: If Y, then don't apply 'Green'.[37] One thing that Travis cases seem to show is that there is no limit to the number of things (conditions/worlds of application) that can be substituted for 'X' and 'Y'. The declarative nature of the substitution rules is problematic because any substitution will be defeasible in light of further evidence, namely, the deployment of a further Travis case. The assumption that rules for application rest on truth conditions creates a rigidity that cannot be sustained in the face of a requirement for a plausible account of learning.

Communication

Suppose that the learning problem is resolved and so all competent speakers have learnt to be able to judge, declaratively, for any set of conditions X, whether or not a predicate applies in that case. Now suppose that one agent says to the other, 'My leaves are green'. What has been communicated? All that the traditional, declarative picture can say is that, assuming the speaker has spoken the truth, the hearer would come to know that one of any possible substitution for 'X' is true. Unfortunately, given that possible substitutions for 'X' are wide, varied and mutually inconsistent, the hearer comes to know nothing. Again, it is the declarative nature of the rules, derived from the idealization that meanings map to truth values, that is causing all the trouble here. There is no way that the hearer can determine that even a coherent subset of the possible substitutions for 'X' is the way the world is being communicated to be. If sentences and predicates encode functions to truth values, then communication is about communicating blunt truths and falsities. Unfortunately, so many truths and falsities are compatible with a function from all possible worlds to a truth value, that communication is impossible. I take this to be a reductio on the traditional model.

Others, such as Cappelen and Lepore (2005) have not taken this to be so. Travis has been accused of being committed to a view that makes communication impossible. Such critics must therefore find a way to explain

[37] Where 'X' and 'Y' are sets of propositions/conditions

away the data from within the confines of the traditional view. I have argued that the prospects for doing this are slim, but this does not prevent the onus from being on me to suggest a substitute for the traditional view compatible with a plausible theory of communication.

2.4 Summary

The orthodoxy I aimed to challenge was that semantics (or a theory of meaning) with no treatment of truth conditions, is not semantics. The job is only partly complete.

The first task was to undermine the claim that truth and truth-conditions can be reasonably appealed to within a semantics or a theory of meaning. That was the goal of this chapter. Traditional, more metaphysical, views on truth, be they inflationary or deflationary seem to be inconsistent with the central inclusion of truth and truth conditions into a theory of meaning. The exception to this rule was logical correspondence. Logical correspondence was designed to marry a theory of truth with a semantics, and Lewis, in making his claim about truth and semantics, came very much from within this discipline. Davidson's project was found to be wanting due to its holism. Montague-Lewis semantic models seemed to be the best bet for incorporating truth into a theory of meaning by constructing a recursive semantics, but such views faced their own problems. By idealising interpretations of terms as functions into truth values, such models end up facing difficulties in accounting for Travis cases. The only way out, for such models, was to appeal to a change in the context of evaluation bringing about a change in the truth value of the proposition expressed. This move is incompatible with pressures derived from learnability and communication.

The second task, and one that must still be completed, is to come up with an alternative suggestion for the interpretation of our terms that is compatible with learning and communication pressures. In this thesis, this task will be restricted to a discussion of predicates. Motivating such an alternative view will be the principal task of the next few chapters.

PART II

COMMUNICATION AND LEARNING

3

Information and Communication

In the last chapter, I argued that truth and truth conditions should not take such a central role in semantics. Over the subsequent two chapters, I will argue that looking at language from a communication perspective motivates viewing statistical correlation as a fundamental notion. In the previous chapter, a loose grasp of communication was all that was needed to undermine the truth conditional model. However, in order to present arguments for an alternative, we will need a theory of communication to build upon.

It is fairly trivial to assert that communication involves the transfer of information. It is less trivial to say what information is, or when one utterance/sentence is more informative than another. These issues will be the main focus of this chapter. Specifically, we will consider (i) what the options are for an account of semantic information, and (ii) how can information be quantified on these accounts. By the end of this chapter, two competing accounts of semantic information will still be on the table. In chapter 4, I will argue in favour of one of them.

In §3.1, I will give a brief introduction to the information theoretic model of communication (also called the *code model*). The notion of communication that will be suggested will be very weak – roughly (and with some extra limitations), the transfer of information from one organism/system to another. Of course, just how weak this conception of communication is will depend on the weakness of the notion of information used. In §3.2, we will consider four bases for informational content: one basis will be truth, another, certainty, and two will be based on be probability. Of the two probabilistic accounts, one will be a relatively novel conception with respect to formal theories of semantic information. Then, in §3.3, we will look at two different models for how semantic informational content can be quantified: One probabilistically, the other in terms of degrees of closeness to the truth.

Finally, in §3.4, we will briefly consider the *inferential model* of communication. The inferential model has been proposed, not as an alternative to the code model, but as an additional component that works alongside it. The inferential model is inspired by Paul Grice's work on implicature and meaning grounded in speakers' intentions (Grice 1989b). In it, inference plays a major role. It will be left open in this chapter to what extent reasoning or inference enters into communication. However, some reasons will be given for why this reasoning/inference needn't be Gricean.

3.1 Communication as the Transference of Information

Based on the work of Shannon (1948), a notion of linguistic communication has been proposed. Shannon's original model describes a framework for communication that is, at first glance, far removed from the sort of model that would fit natural language. It was motivated by the need at the time to better model and develop telecommunications systems; to understand how the structure of the signal, noise in the channel, and the efficiency of coding a message might limit the amount of data transmitted. Information is produced at a source, coded and transmitted, sent as a signal, decoded and received (See Fig. 2). In a noiseless system, by performing the inverse operation of the transmitter, the receiver can retrieve the original message. Noise is a perturbation in the transmitted signal such that the decoding operation performed by the receiver is not guaranteed to output the same signal produced by the information source. Suppose we take the signal to be the audible or legible utterance of some words. What should constitute the transmitter and the receiver, as well as the information source and the destination? Furthermore, in the case of natural language, what might constitute the information that is transmitted?

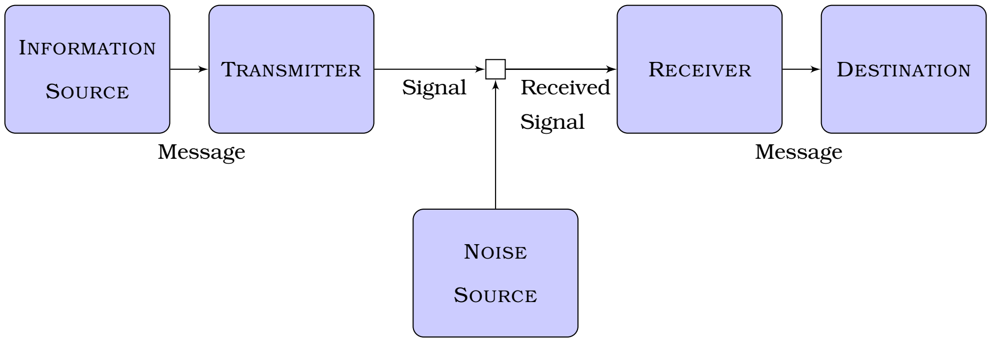

Figure 2 - Schematic Diagram of a General Communication System

Possible answers to the latter question will be addressed in §§3.2–3.3. With respect to the former, the first point to note is that the schema above is highly oversimplified. Very broadly, we might say that speakers constitute both information source and transmitter. Whatever it is that constitutes information, in a noiseless system, speakers produce information via the production of their utterances, thereby transmitting a signal (making noises or making inscriptions). However, there are doubtless further complexities in actual communication. These include the feedback a speaker gets, by way of information from both herself (an internal monitoring of what she is producing), and from her interlocutor (by gestures such as nods and frowns as well as verbal cues of understanding such as, 'uh-huh'). Abstracting away from these complexities, in the following, I shall merely assume that speakers produce signals (encoded messages). These signals can then be heard or seen by another agent and decoded back into whatever information the speaker encoded as the signal. In the rest of the chapter, I will refer to the thing communicated as the *message*, and the means of communicating it as the *signal*. For the purposes of exegesis only, this can be put in some familiar Fregean terminology. Speakers express thoughts (the message) by saying something (producing a signal). Hearers grasp the thought expressed. Put, again for exegetical purposes, in some common-sense psychological terms: speakers believe or know something about the world and can pass on what they know or believe to a hearer by way of a commonly understood language (code).

It should be stressed that the information model of communication does not itself determine what is communicated. Communication is the transference of information, and what we take information to be is dependent on our theory of information. Nonetheless, on any account of information, the basic picture might be held to be too weak. We can take some Gricean inspired cases to demonstrate this. For example, say that we are spying on Jesse with a telescope, but that he is none the wiser. We might discover that Jesse is going to a dance (by observing him putting on his dancing shoes, or by observing him opening an invitation to a dance). By spying on Jesse, observing his actions might be informative in some sense. We might even learn the same had Jesse told us that he was going to a dance, but this does not seem to be a case of communication.

The purpose of this thesis is not to develop a a full-blown account of communication. In the previous chapter, communication featured as the topic of a question designed to test the validity of conceptions of meaning, namely, “Does this conception of meaning make communication possible?”. This question will, in the course of the next two chapters, be turned into a more positive one: “What is communicated?”. On the code model of communication, this question has a simple answer: information is communicated. That means that to answer this question, we should decide what account of information to adopt.

One alternative model of communication that has been proposed (see §3.4 below) is much stronger than the one just outlined. On it, communication is defined in terms of the recognition of speakers' intentions and a rich notion of cooperation (where cooperation is conceived of as requiring reasoning about the mental states of one's interlocutor). This inferential model is stronger than the code model in the sense of being more demanding of speakers, their capacities, and potentially, in there being some rich encoded content over which to reason. There is one simple, methodological, reason for not adopting this kind of inferential model. As mentioned above, inferential model theorists still use the code model. They only claim that some communicative acts require more demanding (inferential) resources from interlocutors themselves. Strictly speaking, this stronger requirement moves the topic from what the words communicate to what, say, thought/proposition etc. is recoverable by interlocutors in context. Our topic is what the information words carry is like (what do words communicate). Given that we can all agree that some encoding and decoding is going on, let us look at the options for what that information is like. Further constraints, generated by inferential concerns, can always be incorporated later.

It should be noted that if one believes that the code model is insufficient to account for all communication, then one needn't be a Gricean. Instead of appealing to cooperation and reasoning about each others beliefs and intentions, some have defended a weaker notion: *coordination* (Barr 2004), (Gregoromichelaki et al. 2011). Coordination is weaker since it does not demand that interlocutors (always) reason over each others' higher-order intentions. Instead, on a coordination account, much of the reasoning we do is egocentric. Notice how weak 'coordination' can be here. Two agents who expressly do not wish to cooperate with one another may still coordinate so as to be able to interact at all. More will be said about cooperation versus coordination in §3.4.

3.2 Defining Semantic Information

If communication is, or at least necessitates, the transfer of information, the question of what (semantic) information is becomes thoroughly central. Shannon's original conception of information was strictly quantitative (see below), and was also not really about individual signals at all, but rather about communication channels, their capacity, and the limits on what efficient coding of signals can achieve. About this *Mathematical Theory of Communication* (MTC), Weaver provided a strict warning:

> "We have to be clear about the rather strange way in which, in this theory, the word 'information' is used; for it has a special sense which, among other things, must not be confused with meaning." (Weaver 1979, p.30)

> "We are concerned not with the meaning of individual messages but with the whole statistical nature of the information source" (Weaver 1979, p. 31)

All that said, any quantitative notion of information does rest upon a less strange conception of information that does tally more closely with meaning, namely, what is communicated. We therefore have a distinction between two types of information: Information (as in content) and information (as in quantity). In §§3.2.1-3.2.4, we will look at a couple of suggestions for the former notion. Here, rather than asking, comparatively, how informative a signal is, we will be asking what signals carry information about. In §3.3, we'll see how different answers to what information (as in content) is, give rise to different answers as to how to calculate information (as in quantity).

3.2.1 Information, Propositions, and Logical Probability

An early attempt to get at semantic information was made by Carnap and Bar-Hillel (Bar-Hillel and Carnap 1952). They treated signals as propositions: things that are either true or false. Relative to what we can think of as possible worlds, signals have a content. So, if our worlds have only two possible facts A and B, each of which either is or is not a fact, there are four possible (complete) *state descriptions* (propositions) that describe each of the four possible worlds: $A \wedge B$, $A \wedge \neg B$, $\neg A \wedge B$, and $\neg A \wedge \neg B$. If our signal is a complete state description, it will be (logically equivalent to) one of the preceding propositions.

The content of a signal is then defined, roughly, as the set of complete state descriptions that are inconsistent with the signal.[1] Of course, messages do not have to be complete state descriptions themselves. The signal/message 'A', will rule out the last two state descriptions. The signal/message '$A \vee B$' will rule out only the last one.

Carnap was interested in a particular kind of probability (what we may call *logical* probability):

> "there are two fundamentally different concepts for which the term *probability* is in general use [...].
>
> (i) Probability$_1$ is the degree of confirmation of a hypothesis h with respect to an evidence statement e, e.g., an observational report. This is a logical, semantical concept. A sentence about this concept is based, not on observation of facts, but on logical analysis; if it is true it is L-true (analytic).
>
> (ii) Probability$_2$ is the relative frequency (in the long run) of one property of events or things with respect to another. A sentence about this concept is factual, empirical."
>
> (Carnap 1962, p. 19)

Logical probability, in the above, is defined as *probability*$_1$. Also mentioned is *probability*$_2$. Over the course of this, and the next chapter, I will be arguing that the central conception of probability underlying semantics is not Carnap's choice of *probability*$_1$, but instead the statistical notion of *probability*$_2$. We needn't accept all of Carnap's distinction when we consider the right kind of information to use as semantic information. In particular,

[1] How this relates to amount of information will be described below.

post-Quine, we needn't accept Carnap's notion of semantics as an investigation into analyticity. Hence, we needn't accept the claim that probability$_1$ is semantic, but probability$_2$ is not. That will be one of the questions under discussion. However, the basic distinction, and how Carnap saw his and Bar-Hillel's project as differing from Shannon's must be borne in mind. Shannon's conception of probability was probability$_2$ (Shannon 1948, p. 5).

Given that Carnap and Bar-Hillel are concerned with probability$_1$, the above definition of content takes on a probabilistic flavour. For example, suppose that our hypothesis is $A \wedge B$ and we are given the information 'A'. That signal rules out two alternative hypotheses ($\neg A \wedge B$ and $\neg A \wedge \neg B$), but it does not confirm our hypothesis, since it does not rule out $A \wedge \neg B$. The probability$_1$ of our hypothesis given the signal and the hypothesis space could therefore be expressed as 0.5. In Carnap's words, "probability$_1$ = degree of confirmation" (Carnap 1962, p 25).

This sort of view of semantic information (as in content) has been summed up as *well-formed, meaningful data* (Floridi 2011, Ch. 4). In other words, a (classical) proposition. Even though information (as in content), so conceived, is propositional, the corresponding information (as in quantity) is based in probabilities. The semantic information (as in content) that a signal carries may be quantified in terms of probabilistic$_1$ confirmation of one's hypothesis.

3.2.2 Information and Certainty

Dretske (1981) was also influenced by Shannon. He develops a notion of semantic information, however Dretske's project was cognitive and epistemic, not strictly speaking linguistic.[2] In this chapter, we will assess Dretske's notion of semantic information, rightly or wrongly, as a candidate for an account of information (as in content, and as in quantity). In chapter 4, we will more closely consider the impact of Dretske's view on the information that words carry.

Unlike Carnap and Bar-Hillel, Dretske defined semantic information partly on the basis of the Shannon inspired notion of information (as in quantity). This introduces a blurring of our distinction between information (as in content) and information (as in quantity). We have been assuming that, in some sense, information (as in quantity) is one way of measuring,

[2]Dretske was interested in sources of intentionality and how we can come to have knowledge. Nonetheless, Dretske frequently uses linguistic examples to make his points. I will do the same in my discussion of his ideas.

for the purposes of comparison, information (as in content). Were we to take Dretske's account of semantic information to be an account of information (as in content), then information (as in quantity) could not be a measure of that. However, Dretske explicitly denies that semantic information carried by a signal is the same as the content of the signal (Dretske 1981, p. 72).[3] This eases matters a little, but we must still try to get clearer on how our information terms are related to Dretske's.

Dretske puts forward a communication condition (A), and two semantic conditions (B) and (C) for a definition of semantic information to satisfy:

> If a signal carries the information that s is F, then it must be the case that:
>
> (A) The signal carries as much information about s as would be generated by s's being F.
>
> ...
>
> (B) s is F.
>
> ...
>
> (C) The quantity of information the signal carries about s is (or includes) that quantity generated by s's being F (and not, say, by s's being G).
>
> (Dretske 1981, pp. 63-4)

Dretske was concerned about how to apply Shannon's formal tools for quantifying information to communication, but he was also worried about the over ambitious application of Shannon's apparatus (pp 53-4). We can assess, roughly, whether some linguistic signal is more or less informative than another (whether some utterance is more or less informative than another), but we lack access to the kinds of probabilities required to calculate these amounts directly. Indeed, this comparative use is central to the account of information that Dretske develops (hence (A) and (C)). Dretske is partially right in making this point. There are probabilities that we probably cannot come to know. However, more plausibly, we may be able to approximate conditional probabilities (as correlations between situations and types of utterances) based on certain idealisations about priors. In chapters 6 and 7, I shall argue that Shannon's apparatus is useful as a MODELLING tool. Even if we can only learn whether some correlations are

[3]Dretske uses the term 'meaning' which he paraphrases as "whatever the person who made it happened to mean or intend by making it".

stronger or weaker than others, this will not prevent us using specific values in a model. In doing so, we will be able to approximate, for example, the quantities of information carried by an utterance, whilst not assuming that these numbers mean anything independently of being greater than or less than others.

Something else that concerned Dretske was truth (hence condition (B)). For him, carrying the information that ϕ is factive in that it necessitates that ϕ. Notice the shift here from probability to necessity. The definition Dretske arrives at combines these considerations about quantity of information carried by a signal, but also that, if a signal carries the information that ϕ, then ϕ. In the following, take k to be the background knowledge of the receiver:

> *Informational content:* A signal r carries the information that s is F := The conditional probability of s's being F, given r (and k), is 1 (but, given k alone, less than 1). (Dretske 1981, p. 65)

Dretske arrives at this result because the conditional probabilities in the definition are not subjective but *nomic* (law-governed). This requirement is, frankly, surprisingly strong to include as part of a conception of semantic information. What it means is that there must be some natural law that connects a signal to the state of affairs described by the message. Put another way, if some signal occurs, and if in virtue of some natural laws (perhaps even natural necessity), this makes some state of affairs occur, then, that the state of affairs occurs is the informational content of the signal. This is, no doubt, a very strong requirement, but we can make more sense of it by realising that Dretske was interested in trying to give an account of an objective notion of information on which to ground an account of coming to know about the world via a decoding of signals. Whether this, in the end is far too strong a requirement will be a central topic of the next chapter. In a system where only $A \wedge B$ can give rise to the message '$A \wedge B$' then the conditional probability $A \wedge B$, given the message '$A \wedge B$' is 1. What this means is that a message carrying the information that ϕ is defined in part as it being the case that ϕ.

It is a little unclear to what extent Dretske's nomic probabilities of 1 relate to Carnap's logical probabilities. Both are supposed to be objective. However, Dretske's are not merely logical. Dretske does allow some nesting of information to be what he calls "analytic" (Dretske 1981, p. 71). For example, any signal that carries the information that something is round also carries the information that it is round or square. However, for him, natural signs can carry information in virtue of nomic nesting: nesting in virtue of natural laws. For example, nested in the information that an accurate bathroom scale reads 80kg is information about the weight of the body upon the scale.

Since a non-subjective, objective conditional probability of 1 entails a truth, Dretske has built truth into the definition of information carrying. Carrying the information that ϕ means, in part, that ϕ. Let us call this position *alethic partiality*. Alethic partiality (partiality with respect to truth value assignment) is a fairly strong notion. On it, carrying information amounts to carrying a true truth-bearer. In contrast to alethic partiality, we have what has been called *alethic neutrality* (Floridi 2011). Alethic neutrality is a weaker position to take with regard to the carrying of information in that a signal may carry the information that ϕ when ϕ is not true.[4] Instead of the meaning of 'a message carrying the information that ϕ', in part, being that ϕ is true (alethic partiality), on an alethically neutral picture, ϕ's truth is seen as being additional to the informational content of the message. The effect of alethic neutrality is that misinformation still counts as information.

Here, then, we have a distinction between Dretske's information based on nomic probability (certainty), and Carnap and Bar-Hillel's logical probability notion. The former is alethically partial: a signal cannot carry the information that ϕ unless ϕ is true. The latter is alethically neutral: false signals can still carry information. Floridi (2011), and by implication, Dretske (1981) take the alethic neutrality of the Bar-Hillel conception of semantic information as a grounds for its rejection. Next, we shall see how something more like Carnap's probability$_2$ could be the basis for an account of semantic information. Recall that probability$_2$ is an empirical conception of probability. On it, probability values represent the relative frequencies of events. Later in §3.3.3, it will be argued that this notion is also free from disputes over alethic neutrality and partiality.

So Dretske's definition of semantic information is meant to capture something about information (as in quantity). That much comes from conditions (A) and (C). What, if anything in Dretske's account, tallies with information (as in content)? Perhaps simply, *that s is F*. Yet that isn't quite right either. For one thing, as mentioned above, Dretske seems to deny this. One example that he gives is meant to draw out the information/meaning distinction:

> "Assuming it to be a law of nature that water expands upon

[4]For an example, see (Fetzer 2004).

> freezing, no signal can carry the information that some body of water is freezing without carrying the information that this body of water is expanding. But the statement, 'This body of water is freezing' can mean that this body of water is freezing without meaning that this body of water is expanding." (Dretske 1981, pp. 72-3)

It is not clear, to me at least, how strongly we should buy into Dretske's intuition. There does seem to be a reading of 'means' in which 'This body of water is freezing' does mean *this body of water is expanding*. Should we say that *this body is expanding* is part of the informational content of the signal? This and other such questions will be a theme for discussion in the next two chapters.

As an added complication, talk of signals, messages and background knowledge generates two readings of Dretske that we must decide between: (i) The information *that s is F* might be carried by the signal, and has a probability of 1 when background knowledge is taken into account. (ii) The signal might not itself carry the information *that s is F*, but *that s is F* could be inferred (with probability 1) from the signal and the background knowledge. I will return to this matter in §3.4.

3.2.3 Information and Truth

Another basis for semantic information is Floridi's combination of Dretske's alethic partiality, and Carnap and Bar-Hillel's conception, with respect to the nature of information (as in content): Well-formed, meaningful, TRUTHFUL data. Put in more familiar terms, a true proposition. The details of this view will be brought out in §3.3 when we consider different approaches to quantifying information.

3.2.4 Information and Uncertainty

As already noted, Shannon's account of information was wholly quantitative. Amongst other things, he was interested in looking at constraints on how, from a set of possible messages, a signal could be coded in such a way as to efficiently and accurately transmit the message across a channel. However, different considerations come into play depending on how probable one message is compared to another. For example, if the sending of the message $A \wedge B$ is vastly more probable (as an empirically witnessed event) than the sending of the message $A \wedge \neg B$, then if the signal '$A \wedge \#$' is received, one is far better off interpreting the message as $A \wedge B$ than as $A \wedge \neg B$.

Similarly to Dretske, and to Carnap and Bar-Hillel, this involves probabilities, too. Such probabilities are not nomic or logical, but *stochastic* or frequency driven. The fact that there are correlations between signals and messages, even imperfect correlations, can be exploited by a system, within some margins of error, to correctly interpret the information being transmitted. This picture is markedly different from the others we have considered. Other accounts of informational content have it that, truly or falsely, signals carry the message that ϕ. On a stochastic account, rarely will informational content be of the form *that* ϕ. Rather, informational content will be related to probable outcomes and stochastic correlations. Over the course of the next few chapters, I will be making the case for this kind of probability as a grounds for semantic information. There are correlations between types of states of affairs and the uses of words (the production of kinds of signals). These imperfect correlations can, I will argue, nonetheless be a fruitful resource that we can exploit to communicate.

In the next chapter, we shall see that, to some extent, this notion of information seems to be what Ruth Millikan describes (Millikan 2000, Appx. B). Millikan adds a condition however, that such correlation must occur for a reason. In chapter 4, arguments will be provided that, as a conception of semantic information, we should adopt something akin to Millikan's weak notion of information, rather than Dretske's nomically grounded one.

3.3 Quantifying Semantic Information

The four bases for semantic information (as in content), namely, truth, certainty, logical, and stochastic probability, can be split into two main categories when it comes to discussing semantic information (as in quantity): Probabilistic and non-probabilistic. Although Dretske's, Bar-Hillel and Carnap's, and my own suggestion for conceptions of semantic information all exploit different notions of probability, the same mathematical modelling of quantities of information can be employed for all three. This mathematics is effectively that of Shannon's Mathematical Theory of Communication (MTC) with some extra refinements and emphasis from Dretske (also present in Bar-Hillel and Carnap). The refinement is required because MTC describes the average amount of information based on the production of a number of possible messages. Quantifying semantic information, on the other hand, will require focusing on specific signals used to describe some aspect of a situation.

Over the next few pages, I will gradually introduce how information is quantified in terms of probabilities. Readers should be aware that at each stage, different notions of probability (logical, nomic, or stochastic) could be applied. Some motivating examples may fit some notions of probability more than others. I leave it up to the reader to substitute their own examples where appropriate. We will first look at the idea that a situation, or at least a type of situation, can contain an amount of information. Then we will consider how someone (or something) with access to this information can use signals to transmit this information to a recipient (a hearer).

Finally, we will look at an alternative way of quantifying information in terms of a measure of distance from the (whole) truth.

3.3.1 Information and Situations

We will use language based examples to demonstrate in what way a quantitative notion of information can relate to certain features of different situations and linguistic messages. The eventual goal will be to see how the telecommunicative vocabulary of Shannon can have an analogue in human communication.

In information theory, the term 'information' is not always used to denote what one might think of as the everyday notion of information. In its technical, information-theoretic sense, information is strictly quantitative. Although this may not seem directly relevant to semantics, it will eventually allow us to account for whether some signal carries more or less information than another. We start with simple situations in which an event occurs. On the information theoretic view, such situations (and events) can be conceived of as themselves containing a certain amount of information. We will see that some events can be more informative than others. Eventually, this will allow us to see a signal as describing a situation or an event. This will have two effects. If signals completely describe a situation, then they will carry more or less information, dependent on the situation that they describe. If two signals describe the same situation, then one signal may carry more information than the other if it more accurately or completely describes that situation.

We will come to signals again later. Here is an example of a simple event. Suppose that a group of eight people must choose one of their number to do a nasty chore.[5] In doing so, they reduce the number of possibilities down to one actuality. We can then ask how much information the situation contains with respect to this decision, or, more succinctly, how informative the decision is.

Carnap and Bar-Hillel propose two different ways that we might try to capture how informative a signal or, in the case above, a decision might be. One way to think about making the decision of who will do the nasty chore is in terms of all the possibilities that it excludes. Suppose that the eight individuals are named Anna, Ben, Charlotte, David, Elena, Fred, George, and Harry. The decision might be made by writing the name of one of the eight on a piece of paper. If 'George' is written, then she and no-one else will have to do the chore (we might think of 'George' as that proposition). The content of 'George' could therefore be assigned as the negation of all the others (first letters are taken for brevity):

$$CONT(\mathtt{G}) = \neg(\mathtt{A} \vee \mathtt{B} \vee \mathtt{C} \vee \mathtt{D} \vee \mathtt{E} \vee \mathtt{F} \vee \mathtt{H})$$

This is similar to the proposal that Carnap and Bar-Hillel make for explicating the intuitive notion of semantic information. In effect, a message is more informative if it eliminates more alternatives.[6]

An alternative way to think about how informative 'George' is, is to think about the the way the group may have chosen George as the unlucky soul. There are eight individuals and so they may have used three separate coin tosses to make a decision. The first toss would narrow the field to four people, the second to two, and the last down to one. In excluding half of the remaining possibilities with each decision, they are working in a binary way. That we assume they made the decision in a binary way is non-essential, but it gives us a measure. Binary information is the number of binary decisions made to reach the main decision. This is Shannon's original choice of measure. For the situation above, the amount of information that this situation contains is three binary units (or *bits*) of information. Had there been only four people to choose from, their decisions would have generated only two bits of information and had there been sixteen present, four units. However, we can also say something about situations where

[5] This example is originally Dretske's.

[6] The Bar-Hillel/Carnap proposal differs slightly. They define a *content element* as the negation of a state description (a full description of a situation/world). CONT is then defined as the class of content elements logically implied by the message.

a decision is left somewhat incomplete. Plausibly, in their coin tossing process, the group may be interrupted after two tosses. This might make it the case that either George or Harry has to do the chore. That would mean that, rather than the full three units being generated, the incomplete decision would generate only two.

Though different, Shannon's, and Carnap and Bar-Hillel's approaches are obviously related. The more possibilities one excludes by a decision, the more binary choices one would need to make to reach it. For the remains of this section, I will stick to the second measure of informativeness (essentially Shannon's measure).

Over the course of the next few pages, we will incrementally add more of Shannon's mathematical tools, and so slowly move closer to a measure that is more suitable for linguistic signals and communication. First, a way of formally determining the information carried by an event will be given. Later, in §3.3.2, we will reintroduce signals (the formal analogues of words being used) and see how they can carry information about a situation, decision, or event.

More can be said about the above kinds of decisions being made in a binary way. In a process such as the one described above, everyone is just as likely to be selected as anyone else. For this kind of special case, the amount of information can be calculated using a base two logarithm:[7]

$$I(s) = \log n \tag{3.1}$$

s	:=	Situation, s
$I(s)$	:=	Information contained in s.
n	:=	Number of possible alternative outcomes in s.

This gives the chore choosing scenario a value of three bits of information ($\log 8 = 3$).[8] However, that says nothing about the details of the situation. Any example of narrowing eight (equally probable) possibilities down to one, will, with respect to that narrowing down, generate three bits of information. Recall though that here, we are only dealing with quantity. Information (as in quantity) is, in a sense, one way to measure information (as in content). In measuring, however, information (as in content) gets stripped

[7]If $x = \log_2 n$, then $2^x = n$. Therefore $\log_2 4 = 2$, $\log_2 8 = 3$ and $\log_2 16 = 4$. Henceforth I will write only "log" which can be assumed to be base two.

[8]The number of possible alternative outcomes in the chore choosing scenario is 8 (one for each person who could be chosen).

away. There is an analogy that can be drawn here with a classical view of semantics. Interpreting the denotations of statements as, say, truth values, also strips them of their content in the sense that two statements with very different contents can share a truth value. Two signals carrying very different information (as in content), can carry the same amount of information (as in quantity).

A different example (which will serve us better later): Suppose that John has one stone in a box and that he has to decide, by flipping a coin, whether to take it out of the box and place it in front of him. There are two possible outcomes, either he ends up with no stones in front of him or with one. Whatever he in fact does, he will have narrowed the possibilities from two down to one. Just like in the above example, John's decision will generate one bit of information because $\log 2 = 1$. With more stones available, the situation can still be modelled as above. Assuming that every number of stones has an equal chance of ending up in front of John[9], the amount of relevant information for this situation can be calculated as in (3.1).[10]

If John flips a coin for every individual stone, then, assuming we are still only interested in the number of stones John ends up with in front of him, things get more complicated as we add more stones to his box. If there are two stones to start with, John would have to flip a coin twice to chose between the stones thus making four possible outcomes, so one might think that the amount of information regarding the number of stones John ends up with is ($\log 4 =$) 2 bits. However, that does not take into account that two of the possible outcomes are indistinguishable. For two coin tosses, there is one outcome in which John has no stones in front of him, one outcome with two stones in front of him, but two possible outcomes for one stone. That means that he has a fifty percent chance of ending up with one stone in front of him.

How much information does John's situation contain? It is here that information theory has a concept to help, *entropy*. Entropy is the measure of the average amount of information contained in a situation. The probabilities of there being no or two stones in front of John are both $\frac{1}{4}$, and the probability of there being one stone gets a probability of $\frac{1}{2}$. The

[9]The more complex case where John decides, for each stone, whether to put it in, is discussed immediately below.

[10]In parallel to the first example, this could be achieved by John using slips of paper with all of the possible numbers of stones written on it and flipping a coin for two equal groups at a time. For example, with eight stones, the first flip would choose between {1, 2, 3, 4} and {5, 6, 7, 8}, the second between, say, {1, 2} and {3, 4} and the third between, say, {3} and {4}.

first step towards finding the entropy of the situation is to establish the amount of information each of these particular results would generate. This information value is known as *surprisal*.

It is this concept that will be central in our application of information theory to words. If there are a number of possible states of affairs/ways for the world to be, and if learning about some of these would be more informative (carry more information) than others, then, if a signal manages to communicate some one rather than another, we might want to say that it, the signal, is more informative (carries more information), than a signal that described some other possibility. We will return to this soon. First, a formal characterisation of surprisal.

The intuitive (but informal) idea behind surprisal is that, if some event is far less probable than another, then learning of that event would be more informative than learning of an event that was anyway known to be highly probable. So, if s_n is the outcome of the situation where n stones are in front of John and $p(s_n)$ is the probability of that result, then for any n, the surprisal of getting that result will stand in an inverse relationship to its probability. Since, there being either no stones or two stones in front of John is less probable than there being one stone, we want the surprisal of there being one stone to be lower than there being two stones (or no stones). The surprisal of something that has a 0.5 probability should be the same as the information generated by a coin toss, namely 1. The surprisal of something that has a 0.25 probability should be the same as the information generated by two coin tosses, namely 2. The surprisal of an event which is certain should be zero and the surprisal of an impossible event should be infinite or at least undefined. This exact measure is given by the log of the reciprocal of the probability. Since $\log \frac{1}{x} = -\log x$, either of the following will give the identical result:

$$\begin{aligned} I(s_n) &= \log \frac{1}{p(s_n)} \\ &= -\log p(s_n) \end{aligned} \tag{3.2}$$

s_n	:=	Situation, s, where n stones are selected.
$p(s_n)$	:=	Probability of s_n.

The thought behind (3.2) is that surprisal is the inverse of probability – the more probable something is, the less surprised you should be and vice versa.[11] In that sense, the more improbable an event is, the more informative it is. The values for $\log x$, when $x = 1$ and upwards, increase in ever smaller amounts (the opposite of a graph for 2^x which increases in ever greater amounts). For values of x less than 1, $\log x$ is always negative and becomes increasingly so. Because probabilities are always less than or equal to 1, $\log \frac{1}{p(s_n)}$ will increase as $p(s_n)$ gets smaller (an impossible event would be infinitely surprising) and will be close to 0 when $p(s_n)$ is close to 1.

To recap slightly, $I(s)$ in (3.1) was the information contained within a special kind of situation (where all eventualities are equally probable). It should, however, be thought of more as a situation type. Given no particular choice of number of stones, the situation can nonetheless contain some average amount of information. Within a situation type, there can be multiple possibilities (no stones, 1 stone, 2 stones), and surprisal ($I(s_n)$) is a measure of how much information that one result in particular would give. In that sense it is the information given by a token of a situation type. In the John example, there are n possibilities and so $I(s_n)$ is the information given by some instantiation of the situation type where John puts n stones in front of him. When all possibilities are equally probable, (3.1) gives the information contained within a situation type. When they aren't, you can take a weighted average of surprisal values using the equation (3.3) given below.

Given the probabilities in the two stones case, zero stones would have a surprisal of 2, as would two stones. One stone would have a surprisal value of 1. The average information for the situation can then be given as the sum of these surprisal values weighted against their probability:

$$\begin{aligned} I(s) &= \sum_n p(s_n) \times I(s_n) \qquad (3.3)\\ &= p(s_0) \times I(s_0) \quad + \quad p(s_1) \times I(s_1) \quad + \quad p(s_2) \times I(s_2) \\ &= 0.25 \times 2 \quad + \quad 0.5 \times 1 \quad + \quad 0.25 \times 2 \\ &= 1.5 \end{aligned}$$

If we added an extra stone to the situation, there would be eight variations, but only four outcomes: $p(s_0) = p(s_3) = \frac{1}{8}$ and $p(s_1) = p(s_2) = \frac{3}{8}$. This would make the entropy of the situation with three stones approximately

[11] That is the intuitive notion of surprisal. It should be stressed that 'surprisal', as a technical term does not need to be interpreted in such psychological terms.

1.81. This tells us that, on average, learning the number of stones John ends up with in front of him amounts to learning more information the higher the number of stones he starts with. This is the result we would expect because more possibilities are being introduced with every extra stone in the situation.

Dretske (1981, p. 48) pointed out that, even though entropy is a useful notion in telecommunications, it is less useful as a semantic notion. Suppose that once he has sorted his stones out, John tells Mary how many are in front of him. It is less interesting to know how much information Mary will find out on average[12] than it is to know how much information a particular result would convey. Therefore, as suggested earlier, the main notion we should be looking at is surprisal not entropy. Dretske's suggestion that surprisal is more important for semantic information than entropy is a move away from Shannon, but remains similar to Carnap and Bar-Hillel's definition. In plain terms, we are interested in what information a single decision carries about a situation, not what the average amount of information that a situation containing a number of possible decisions contains.

3.3.2 From Decisions to Signals

We are now in the position to formally describe how much information one of a possible number of events carries in a given situation. The next step will be to bring in signals. Signals carry messages, and so if a signal carries the message that one of a possible number of events in a given situation has occurred, then we will want to be able to say how much information about that situation the signal carries. The basic picture for this will be fairly simple. If a signal indicates that an event has occurred (and no more or no less), then the signal can be said to carry the same amount of information as the event itself. Of course language is rarely going to be so neat and tidy. That means that we will also need to be able to describe a signal carrying more/too much information, as well as one carrying less/too little information.

We can now add to the example of John and his stones. The same process will take place as before where some number of stones is selected to be placed in front of John. Now, however, John will tell Mary (signal to her) how many stones are in front of him. To capture how accurately John's signal

[12]This would be akin to being told the average amount of information carried by John's signal, were he to repeat the process many times and tell Mary the result each time.

conveys the situation to Mary, three more information theoretic notions are needed. The first is simply a measure of the amount of information that is transmitted from a source to a receiver. If John's stone sorting is the source and his passing a slip of paper with a number on it is the transmission of a signal, Mary, as a receiver, can get a certain amount of information from the source via the message.[13] If all is optimal, the amount of information gained is equal to that generated by the source. Where $I_s(r)$ is a measure of this transfer, in the optimal case, $I(s) = I_s(r)$.

$$I_s(r) \quad := \quad \text{Information about } s \text{ received at } r.$$

The two other notions concern cases where the amount of information that a signal conveys about a situation (source) differs from the amount of information produced by the source (when $I(s) \neq I_s(r)$). Sometimes, a message may fail to convey some information that is contained within a situation. The amount of information contained in a situation that is not conveyed by a message is called the *equivocation*. With or without equivocation, sometimes the message itself contains more information than is contained within the situation. This amount of extra information is referred to as *noise*. It is worth noting that the everyday term 'noise' tends to cover both of these two notions. When an analogue radio hisses, not only does the hiss add some information to the receiver's situation (adding noise in the technical sense), it may also obscure some of the radio message itself thereby reducing the amount of information received from the source (adding equivocation). Digital radios, on the other hand, tend to leave gaps of silence when they lose signal. Assuming that, in this context, silence is uninformative, one could truly say that digital radios are noiseless. Unfortunately, what we arguably most often care about is equivocation. We care about receiving all the information from the source, even if not only that information. The digital switch-over, if and when it comes, will not help us in that respect.

For example, in the case where John starts with eight stones, where every number has an equal chance of ending up in front of him, we said that this situation contains three bits of information. Say John ends up with four stones, so writes the word "four" on a slip of paper. Unfortunately a drop of rain hits the paper and so only "f ###" can be read from the

[13] The communicative exchange is oversimplified.

message. Mary knows that John started with eight stones and so when she receives the note, not all of the information is accessible to her. For Mary, since "f###" could be "four" or "five", eight possibilities are narrowed down to two.[14] $I(s) = 3$, but one bit of information is lost ($I_s(r) = 2$).[15] This means that the equivocation is equal to one bit. The relation between equivocation, the source information, and the amount of information from the source to the receiver holds in the following way:

$$I_s(r) = I(s) - \textit{equivocation} \tag{3.4}$$

However, in the above case, we could look at the amount of information in Mary's situation: $I(r)$. She receives a slip of paper with "f###" which contains at least two bits of information because it narrows eight possible situations down to two. However, the paper contains more information than that. For example, that there was water about when the note was transferred. This extra information is not considered to be information generated by the source (which we restricted to only the number of stones John ends up with), and so we could also say that however much extra information is transmitted by the message, the information transmitted from the source only is equal to the information in the receiver's situation minus all the other (irrelevant) information, or noise, in that situation:

$$I_s(r) = I(r) - \textit{noise} \tag{3.5}$$

Again, as Dretske points out, for semantic purposes, we are not interested in the average amount of information about a source transmitted by a signal, but the amount of information about a source transmitted by a particular signal with respect to a particular situation. Therefore, we do not need to concern ourselves with the more complex formulas that Shannon introduces for determining the average amount of information, given an average amount of noise or equivocation generated. However, we

[14] As an aside, it is worth mentioning another of Shannon's notions, *redundancy*. Imagine that the signal was "fi##". Mary could still recover *five stones were selected* as the message. If we see each of the letters used in the signal as information carriers, this suggests that some of the information carried is redundant. One of Shannon's major results was to show how redundancy is actually useful for combating noise and equivocation.

[15] This result will be shown formally below.

can look at one further equation that will give a formal basis for what we have already done informally. This equation, suggested by Dretske, will allow us to calculate the equivocation of a signal based on the number of possible situations it leaves open.

Think of the previous example where the message gets blurred. There were eight possibilities, but when "four" was written down one couldn't tell if it was five or four stones that were selected. Equivocation is a measure of how much information is lost from a message and one way of starting to get a handle on that is by a conditional probability. We can ask what the probability is of a situation, given some particular signal. If, as before, s_n is a situation in which n stones were chosen, and r_m is the situation where Mary has received a signal m, then a conditional probability can be given for the case where the signal was "f###": $p(s_n|r_{f\#})$. With that formula we can find the probability of every situation given the signal 'f###'. However, that is not enough. For all non-zero possibilities, we also want to get a measure on the surprisal of that result (i.e. of there being n stones, given the signal 'f###'). The surprisal values can then be weighted according to their probabilities and summed:

$$E_{f\#} = -\sum_n p(s_n|r_{f\#}) \times \log p(s_n|r_{f\#}) \tag{3.6}$$

$E_{f\#}$	:=	Equivocation of the signal 'f###'.
$r_{f\#}$	:=	Situation, r in which signal 'f###' is received.
$p(s_n\|r_{f\#})$	:=	Conditional probability of s_n, given $r_{f\#}$

For the above case there are only two non-zero results to be summed. This is because, when the signal "f###" is received, the only equivocation occurs between there being four stones or five stones:

$$\begin{aligned} E_{f\#} &= p(s_4|r_{f\#}) \times \log p(s_4|r_{f\#}) \quad + \quad p(s_5|r_{f\#}) \times \log p(s_5|r_{f\#}) \\ &= 0.5 \times 1 \quad + \quad 0.5 \times 1 \\ &= 1 \end{aligned} \tag{3.7}$$

What this says is that when Mary sees the signal "f###" and, originally, the message was *there are four stones*, since there is equal probability that the message was *there are five stones*, given what the signal legibly says, she can no longer distinguish between these two possible messages. This amounts to the loss of one bit of information and that is exactly the result that was anticipated a few paragraphs previously.

In sum: We now have a way of quantifying the probabilistic information (as in content) carried by a signal (with respect to some situation). We also have a way of expressing how information can be lost in transmission (even if not in translation). However, we shouldn't focus on the amounts of information carried by a signal independently of any other signal. Amounts of information will be comparative only. They will allow us to compare, for example, whether one utterance would be more or less informative than another. In chapter 7, these tools will be applied to vague expressions. There we will use an informativeness measure to model formal pragmatics. In particular, information (as in quantity) will allow a more exact understanding of Grice's maxim of quantity.

3.3.3 Information and Alethic Neutrality

Recall that Dretske's and Floridi's accounts of information are alethically partial (if a signal carries the information that ϕ, then ϕ). But also recall that Carnap and Bar-Hillel's theory is alethically neutral (if a signal carries the information that ϕ, then ϕ or $\neg\phi$). Bar-Hillel and Carnap deny that semantic information implies truth:

> "It might perhaps, at first, seem strange that a self-contradictory sentence, hence one which no ideal receiver would accept, is regarded as carrying with it the most inclusive information. It should, however, be emphasised that semantic information is not meant as implying truth. A false sentence which happens to say much is thereby highly informative in our sense." (Bar-Hillel and Carnap 1952, pp 7-8)

Both Dretske (1981, p. 44) and Floridi (2011, Ch. 4) argue that semantic information cannot be alethically neutral. If they are right and alethic neutrality is untenable, then Bar-Hillel and Carnap's conception of information is threatened. Fortunately, our purposes do not require that arguments about alethic neutrality be settled. In the next chapter, I shall argue that semantic information should be modelled as something close to

Shannon's original conception. It needs to be based on probabilities grounded in statistical frequencies. As a part of this argument, I will contend that Dretske's notion is too strong for communication and cannot apply to natural language.

The reason why alethic neutrality should not concern us is that it does not make sense to ask of a statistically based notion of information, whether it is alethically neutral or partial. Recall that on this conception, the amount of information that a signal carries is inversely proportional to the probability of some situation, given a signal. If this probability is merely a reflection of statistical frequency, then the information the signal carries is not the kind of thing that can be associated with truth values. For example, suppose that the statistical frequency grounded probability of a situation s given a signal r is 0.9. That means that the probability that not-s, given r is 0.1. Even if, on the basis of r, it would, given past correlations, be highly reasonable to conclude that s will arise, the information carried by r also conveys that this is not always the case. Were we, misleadingly, to say that the content of r is s, the information carried by r incorporates the fact that sometimes occurrences of r are truthful and sometimes not. More accurately, the informational content of r reflects how reliable r is as a signal of s. This takes into account the possibility of not-s, given r. It therefore makes little sense to ask if the transference of the information carried by r requires it to be true that s arises. More generally, if r says/conveys the extent to which s is probable, where 'probable' simply conveys a correlation between instances of r and instances of s, then what r conveys is not the sort of thing to be true or false.

This inclusion of reliability, where a signal is a more reliable sign if it is more strongly correlated with a state of affairs, will be of importance in the next chapter where we consider what information has to be like in order for it to be useful.

3.3.4 Verisimilitude and Floridi's Alternative

Before proceeding down a probabilistic route, there is at least one major alternative to consider. Instead of taking conditional probabilities as the basis of a measure, one can instead assess statements as being closer or further from the truth. In fact, on this approach, one gets two possible measures. Suppose, similarly to the view proposed by Bar-Hillel and Carnap, that a situation is defined as consisting of a set of facts. Statements may convey some of those facts, but be silent on others. Other statements may similarly express different numbers of facts, ranging from all to very few. None of these statements will be false, but some will fit the facts more fully than others, and so a measure of truth-likeness may be useful in comparing these statements. As well as all of the true statements about a situation, one might be able to say many false things about that situation. Again, some of these statements may be better than others. Whereas some may express the negation of all of the facts of the situation, others may get a lot right, but some wrong. A measure of the truth-likeness of these false statements could be useful in comparing them.

For example, suppose that a waiter is taking the order of *Table-Five*. The order is one mushroom risotto and one fiorentina pizza. The waiter, unwisely given the temperament of the chef, could tell the kitchen, "Table-Five have ordered a mushroom risotto". This statement is true, but it fits the facts less well than "Table-Five have ordered one mushroom risotto and one pizza", and that statement fits that facts less well than "Table-Five have ordered one mushroom risotto and one fiorentina pizza". Alternatively, the waiter might, perhaps even more unwisely, say something false such as, "Table-Five have ordered one mushroom risotto, one fiorentina pizza, and a garlic bread'. But that wouldn't be as bad as saying, '"Table five have ordered two mushroom risottos, two fiorentina pizzas, and a garlic bread".

Floridi (Floridi 2011, ch 5.) introduces the terms 'vacuity' and 'inaccuracy': The true statements contain vacuous information. Statements can be true, but with varying levels of *vacuity*. The false cases are inaccurate and contain various levels of *inaccuracy*. Vacuous statements contain more information than was contained within the situation, and inaccurate statements contain less. In that sense vacuity and inaccuracy are the truth-likeness analogues for equivocation and noise, respectively. Noise is added information, so all noise is, in a sense, vacuous. Equivocation is lost information. So equivocation might lead to inaccuracy. The two pairs are not, however, cognates. Only true statements can be vacuous and only false statements can be inaccurate. Given non-contradiction, that means that no statement can be both vacuous and inaccurate. The same is not true of noise and equivocation. Sometimes, on the phone, instead of hearing the other person, all you hear is an echo of your own voice. Put in information theoretic terms, all of the information carried by the signal of your interlocutor has been lost (equivocation is maximal), and the echo is just noise.

Floridi's distinction is reminiscent of work on truth-likeness or *verisimil-*

itude which started with Popper (Popper 1963) and saw a renaissance in the seventies and eighties (for an overview see (Niiniluoto 1978)). Floridi essentially applies some of these ideas to the philosophy of information. He proposes a non-probabilistic approach to get a measure on the informativeness of a statement, derived from degrees of vacuity and inaccuracy. The account is non-probabilistic, since degrees of vacuity/inaccuracy are not probabilities of anything, but measures of something like closeness to (complete) truth and closeness to (complete) falsity. Floridi's work is also of some interest, given that he makes a suggestion for a treatment of vague statements.

The first part of the account bears some resemblance to Bar-Hillel and Carnap's CONT. Whereas Bar-Hillel and Carnap define the content of a signal in terms of the number of complete state descriptions incompatible with it, Floridi describes signals as *infons*[16] which are supported by some situation to different degrees. If supported by this situation, Floridi calls the infon 'true'. This is a departure from how infons are normally understood. For example, Devlin (2006, pp. 603) is explicit that infons are not the sorts of things that can be true or false. In situation theory, what is true or false is a proposition which says of a situation that it supports an infon (contains some bit of information). This is a non-trivial departure that Floridi is making. He also wishes to define degrees of closeness between infons and a "state of the world" (Floridi 2011, p. 118), where states of the world are given as strings of conjoined propositions. This really begins to sound as though Floridi's infons are true propositions and situations are sets of propositions, namely, possible worlds.[17]

This makes Floridi's notion of content very close to the orthodoxy view discussed in the last chapter. Messages are (possibly complex) propositions in the sense that their content is definable as their truth-value evaluated at an index. The only difference between Floridi's account of alethically partial information and the orthodoxy is that, amongst the true statements, we can say which give a fuller account of all the facts of the situation, and amongst the false statements, which gives the account closest to the truth.

With respect to vagueness, this makes Floridi's position epistemic (Floridi 2011, p. 120). He assigns the vagueness of signals as being uncertainty over the degree of accuracy or vacuity that the signal has with respect to the

[16]The term 'infon' comes from situation theory and situation semantics (See (Devlin 2006) for an overview). I will use infons and other situation theoretic notions in chapter 6.

[17]As pointed out to me by Robin Cooper, one could read Floridi's 'true infon' as an infon which is supported by a situation. Nonetheless, the spirit of Floridi's account still seems to be alien to the more local approach found in situation theory.

described situation. There is a promising idea here, one that will be followed up in later chapters, that the information a vague expression carries is mismatched in some way with the situation it is used to describe. However, due to the way Floridi ties truth to information, given the arguments presented in chapter 2, his notion of information must be rejected.

3.4 Code vs. Inference

At the outset of Chapter 1, it was remarked upon that there is a good deal of division in the philosophy of language over the kinds of things that are vague. Amongst the options were sentences, utterances, propositions and beliefs/mental states. Similarly, we might ask by what is linguistic communication done? Dretske's and Floridi's accounts are compatible with the idea that (true) thoughts/propositions are communicated. For them information is either something that conveys certainty (adds to knowledge), or that conveys truth. Reading Dretske as being similar to Floridi would be compatible with reading information (as in content), as that entity which is carried by a signal and is true (nomically certain) when all background knowledge is taken into account. This is the first reading of Dretske mentioned earlier. The weaker conception of Carnap and Bar-Hillel is also compatible with the idea that thoughts/propositions are communicated. For them, however, false thoughts can also count as conveying information. The alternative, Shannon inspired account has what is transmitted by a message in communication as being much weaker: something more like an indicator of probable states of affairs.

If we took the line that thoughts (true or false) are what are communicated by words (a view this thesis will slowly try to dislodge), then maybe sentences communicate thoughts. Alternatively, we might think that it's the use of words (by people producing utterances) that linguistically communicate thoughts.[18] Since Austin (and the later Wittgenstein), via, amongst many others, Searle, Strawson, Grice, Kaplan, Putnam and their descendants, the idea that sentences communicate thoughts has been seen to be thoroughly wrong. At the very least, the presence of indexical pronouns and tense markers ensures this result. However, even if we stick to uses of words, by individuals in some situation, there are still a great many puzzles

[18]Mental states and beliefs might feature somewhere in the story about communication (see below), perhaps as the things communicated, but it isn't beliefs that communicate thoughts. Equally, propositions might be held to be what gets communicated, but propositions can hardly be the things that communicate thoughts.

about how individuals come to express and grasp thoughts.

Sperber and Wilson (1986) argue against what they saw as the standard assumption about how communication works. On that assumption, communication is achieved by encoding and decoding messages. If coding and decoding is all there is to communication, but sentences do not encode thoughts (propositions), then some other kind of information must be being decoded. One proposed fix for the code model is that one needs both sentences uttered and mutually known propositions to decode thoughts. Mutual knowledge (Lewis 1969) is the class of propositions that interlocutors know, and know that each other know etc. Sperber and Wilson found the code-model fix to be psychologically implausible. I will not go into the details of their objections here, but I will briefly summarise their alternative.

Relevance Theory suggests that instead of merely explaining communication in terms of code, another separate and irreducible form of communication is inference. Sperber and Wilson weakened the common knowledge hypothesis to one of *mutual manifestness*. The idea is that manifest facts are those things perceivable and/or inferable in our environment. Mutually manifest facts are those that are manifest to interlocutors, and manifest that it is manifest to each other, again, *etc.*. Linguistic communication is a matter of using what is mutually manifest, and linguistically encoded information, to infer what the speaker intends to convey.

An analysis of this kind of picture will not be undertaken here, but I will make a few comments. First, this notion of communication is much stronger than the weak one outlined above. The simple informational model appealed to coordination which, it has been argued (Barr 2004), (Gregoromichelaki et al. 2011) can be emergent from interaction between non-cooperative agents. In particular, no appeal to mutual knowledge or even mutual manifestness needs to be made. However, even if it turns out that we do require a model of communication via intention recognition, we will not assume such a strong model in any of the following, where different notions of what is communicated (in terms of information) are discussed. This is acceptable since, even on an inferential model, linguistic signals still encode information. The claim we need not accept is that the message communicated is only recoverable via a process of reasoning about the (higher-order) intentions of the speaker. Second, even if we accept that a good deal of inference goes on in communication, even that some communication is done by inference alone, there are good reasons not to assume a neo-Gricean model. Critics of this model come from a variety of disciplines. In philosophy, (Millikan 1984, ch. 3), (Millikan 2005b, ch. 10), (Travis 2008c), (Travis 2008a). In linguistics, (Gregoromichelaki et al. 2011). In psychology, for an overview see (Keysar et al. 1998). For our purposes, entering into such disputes would be too much of a diversion, and it would be well outside of the scope of this thesis to even begin to try to develop a substantial model of communication. For now, let us simply be open to the idea that inference, or at least reasoning, goes on in communication, as well as some decoding of signals[19]. However, let us also be sceptical towards the claim that understanding others necessitates reasoning about their mental states.

3.4.1 Codes, Inference and Truth: Chapter 2 Revisited

Dretske

In their discussion of the code model, one of the authors Sperber and Wilson seem to have in mind is Dretske. In the next chapter we'll see how Dretske, in the end, concludes that linguistic signals do not carry information.[20] However, at the risk of building a straw man, we can using some of the insights from Sperber and Wilson to evaluate whether Dretskean information, viewed as linguistic, is vulnerable to the arguments from the end of chapter 2.

This brings us back to the two readings of Dretske (1981) mentioned in §3.2.2. I take these in turn. The first version of Dretske was such that the utterance a speaker produces carries the information that ϕ and, given the signal and given some amount of background knowledge, ϕ receives a nomic probability of 1, (but would be less than 1 without the signal). This reading of Dretske, is directly incompatible with the arguments from chapter 2. It amounts to the model where an utterance expresses some proposition, and has a single value relative to a context.

The other reading holds that utterances do not express propositions *per se*, but that the proposition expressed can be inferred from the utterance and the background knowledge (and has a probability of 1). This is not necessarily any kind of deductive inference, but some form of probabilistic reasoning from a signal and some background knowledge to the certainty of

[19]Even Sperber and Wilson think both processes are present, and so this is uncontroversial.

[20]This conclusion is based on the fact that words can be used to misrepresent. Given the alethic partiality of Dretske's notion of misrepresentation, any kind of misinforming is ruled out by definition.

a proposition.

This is a different picture from the one we had at the end of chapter 2. On this reading of Dretske, words (the signal) carry some data. Whatever these data are, they interact with some background knowledge, and their combination gives ϕ a probability of 1. That means that the signal itself need not, independent of the right background knowledge, carry any propositional information. It also means, though, that what the words themselves carry might be far simpler to learn and far simpler to grasp. This suggests that Dretske's information (as in content) might not succumb to learning and communication problems in the same way.[21]

However, the presence of nomic certainty on this reading makes for an alternative problem. Nomic certainty, recall, is a certainty grounded in natural laws. If the data carried by the signal (what we might call the meaning of the signal) do not on their own make anything nomically certain, but do make ϕ nomically certain when considered in conjunction with some background knowledge k, then the signal carries the information that ϕ. However, nomic certainty, once attained as a conditional probability, can never be withdrawn:

If:

$$p(\phi|k) = 1$$

Then:

$$p(\phi|k, k') = 1 \quad \text{for all } k'$$

$$\begin{array}{lll} \phi & := & \text{The message, } \phi. \\ k, k' & := & \text{Sets of background knowledge.} \end{array}$$

Because nomic certainty implies truth, if a signal says that/carries the information that ϕ, given some circumstance described by k, then there simply cannot be any addition to those circumstances which would mean that ϕ were less than nomically certain. So ϕ could not fail to be anything other than true. This flies in the face of examples from chapter 2 in which changing circumstances do seem to make for changes in truth value. This in itself suggests that something weaker might be needed.

[21] It should be borne in mind here that reading Dretske this way does not require that we buy into a mutual knowledge (or a mutual manifestness) thesis. It may simply be enough for an agent to know something, irrespective of whether she knows that her interlocutor does.

However, we will still return to Dretske in the next chapter. That is because we must still evaluate his account of how (semantic) representation is related to information.

Carnap and Bar-Hillel

On the basis of chapter 2, Carnap and Bar-Hillel's conception also faces a challenge. Recall that for them, the information (as in content) of a signal can be given as a class of negated state descriptions. Such a view of content makes their account seem like a kind of proto-dynamic update semantics wherein propositions update the context (the set of worlds), by intersecting it. At any rate, on this view, for every signal, there is a class of propositions that are incompatible with that signal. After all, bear in mind that information (as in content) is, relative to some actual complete state description, still supposed to be either true or false in virtue of the content it has. This is highly problematic given the considerations of chapter 2. We had reasons to reject Floridi's conception. Here similar reasons count against Carnap and Bar-Hillel's account.

3.5 Summary

One purpose of this chapter was to outline a basic model of communication that will serve a specific purpose. It will act as a guide for the kind of picture of meaning that we should adopt. Applying this to linguistic communication then raised the question of what semantic information is (what is being transferred), and, when is some utterance (signal) more informative than another (how much is being transferred). The model is only as strong or as weak as the notion of information (as in content) that is plugged into it. The proposed model was therefore very loose, essentially just a requirement that information be transferred between agents. If it turns out that our notion of information (as in content) cannot plausibly be something systematically transferred between individuals, then this will be reason for not adopting that notion of information.

However, we saw that some suggestions for information (as in content) in the literature are just those suggestions we considered as candidates for meaning in chapter 2. If 'information' just turns out to be a euphemism for 'proposition' or, even more restrictively, 'true proposition', then any case we had for the implausibility of communicating only (true) propositions will be just as strong for the implausibility of communicating/transferring that kind of information. On this basis, we ended up rejecting both Floridi's, and Carnap and Bar-Hillel's accounts of semantic information.

We also rejected readings of (Dretske 1981) on which his account of information is straightforwardly an account of the information carried by words. However, Dretske will be taken up again in the next chapter. Dretske argues that utterances and beliefs frequently fail to carry information. I will provide reasons to reject this conclusion. This will involve providing details for the statistical account of semantic information that was outlined in this chapter. This alternative, much weaker conception of information was defined in terms of frequencies. The information (as in content and as in quantity) that a sign carries is determined by how reliable that sign is as an indicator of certain states of affairs. On this account, rarely, if ever, will a sign carry the information that ϕ where 'ϕ' is any kind of orthodox proposition.

In chapter 5, we return to learning, where it will be argued that the *semantic bottleneck* problem calls for learners to associate words with defeasible generalisations that can be integrated with the context. There it will be suggested that a plausible model for such generalisations is one related to the stochastic proposal for semantic information.

4

Semantic Information and Correlation

In the last chapter, a very basic outline for an account of semantic information was given. This account is based on statistical correlations. Other accounts were rejected, but still to be discussed are the implications of Dretske's view on linguistic signs and a later version of his account of information.

I will describe how Dretske's views on information lead him to conclude that beliefs (what he takes to be bearers of content) frequently fail to carry information. The same point could be applied to utterances (if they are bearers of content), in which case, one should conclude that utterances frequently fail to carry information. The main purpose of this chapter should be seen as an attempt to (i) reject this move, and (ii) propose an alternative. I will address (i) in §§4.1-4.2 and (ii) in §4.3.

In §4.1.1, I detail how Dretske's view of natural information changed in *Explaining Behaviour* (Dretske 1988). In §4.1.2, I will detail and criticise Dretske's account of linguistic representation. In §4.2, I will use and adapt some arguments of Ruth Millikan's to criticise Dretske's notion of information as applying to either natural or linguistic signs.

In §4.3, we will again look to Millikan. We will consider Millikan's statistical frequency based account of natural information and her claim that the only place where laws need to enter an account of natural information is to explain why statistical frequencies persist. I will put forward the view that it is here that we should define the distinction between natural and linguistic (non-natural) signs. Linguistic signs carry information grounded in correlations, but they persist for different reasons. I will claim that, although laws do not explain the persistence of linguistic correlations, they still persist for good reason. These good reasons will be spelled out via appeal to Millikan's account of linguistic convention.

4.1 Information for Communication

Naturalising intentional signs is a long-standing project within philosophy, and accounts of information seem to provide a way to achieve this goal. If a naturalistically acceptable account of how natural signs carry information can be given, then perhaps a naturalistically acceptable account of how other signs (such as linguistic signs) carry information, might also be available. Dretske was a pioneer in philosophy for formally trying to realise this project. As we saw in the last chapter, Dretske's 1981 account grounds information in natural laws (nomic certainties). Roughly, if there are some laws that ensure a perfect correlation between a sign and a state of affairs, then that sign carries the information that the state of affairs obtains. A classic example is the rings in the trunk of a tree. These rings carry information not only about the age of the tree, but also about things such as the amount of water in the environment during that year of growth. Despite this, the rings on the tree are neither true nor false. This is not the case with an appropriately situated utterance of 'There are 5 rings on that tree'. Natural signs such as tree rings cannot misinform or falsely inform. Utterances of 'There are 5 rings on that tree' can.

Dretske (1981) was more interested in beliefs than utterances, but a similar disanalogy holds. Beliefs can, seemingly, have content and be true or false. Natural signs can have content (in the sense of carry information), but certainly cannot be false.[1] This creates the problem that beliefs/utterances cannot carry information when they are false, even though they apparently have the same content as when they are true. His solution (Dretske 1981, ch. 8) is to see beliefs (in general) as having information carrying roles, however he claims that INSTANCES of beliefs can fail to carry the information appropriate to their role. That means that sometimes beliefs can have meaning but not carry information. An analogous move could be made with

[1]One might, however, claim that the information that natural signs carry is true.

respect to linguistic signs (utterances). Words, or maybe utterance types,[2] could have an information carrying role, but nonetheless fail to carry that information on instances of being uttered.

Clearly, however, this line of thought is motivated less by clear distinctions between natural signs and linguistic signs, than it is by the idea that what natural signs indicate is in some sense true, whereas what linguistic signs indicate is not always so. Put another way, this line of thought only gets going if one accepts a strong view of (natural) information such as Dretske's. This conception of information will be targeted in §4.2.

4.1.1 Dretske, Nomic Certainties, and Statistical Frequencies

In later work, Dretske's view of natural information changed slightly, and his way of describing information does too. In the language of (Dretske 1988), the sign *indicates* the state of affairs. For example, the number of rings might indicate the age of the trees, even though we cannot sensibly ask if the rings in the tree are true. Indication is meant to contrast with (mere) representation in a sense which is essentially that of Grice's *natural meaning* and *non-natural meaning* (Grice 1957). Dretske's notion of indication is also meant to tally closely with his notion of information in (Dretske 1981). There are some differences however. First, indication is not meant to be dependent on an agent. Recall that Dretske's definition of semantic information included a reference to background knowledge. Dretske (1988) removes this reference to knowledge. A natural sign indicates some state of affairs if the state of affairs is certain, given the sign, in virtue of some natural law. Second is a point noticed by Millikan. In Dretske (1988), he seems to entertain the notion of a weaker form of information/indication that is not purely nomically governed.

Dretske's weaker form of indication allows that, in some cases, statistical frequencies can affect whether or not a signal carries information about (indicates) some state of affairs. Millikan notices that this claim can be read in different ways. To see how, we need to understand how even natural signs can go wrong. An example that Dretske uses is a petrol gauge. Given the laws that govern the workings of the gauge, we are told that the gauge being in a certain position (pointing at $\frac{1}{2}$, say) indicates that the petrol tank (in that car) is half full. However, if certain things are done to the gauge or its environment, then the gauge pointing where it does will no longer indicate

[2] Cf. The discussion in (Dretske 1981, pp. 190-3) on maps.

that state of affairs. For example, if we filled the tank with lead weights, or connected the gauge to another tank in another car, it would not indicate that the petrol tank (in the car the gauge is in) is half full. Dretske explains this as the petrol gauge not having the right *channel conditions*. So, even though the gauge indicates that the tank is half full (with nomic certainty), it does so only relative to the right channel conditions. It is possible, Millikan notes, that the right channel conditions rarely hold (that many petrol tanks are not accurate is her case in point). But also, how probable it is that they hold need be no more than a statistical frequency. Hence, indication (carrying natural information) might be understood as states of affairs being nomically certain, given some merely statistically probable set of background conditions. This is weaker than Dretske's original definition which does not refer to the probability of channel conditions holding.

Further, Millikan noticed that frequencies enter into Dretske's account at the source situation too. Dretske seems to allow that what a sign indicates can be affected by (merely) statistically probable conditions in the source situation. Dretske says that doorbells ringing mean (as in indicate) that somebody is at the door. However:

> "If squirrels changed their habits (because, say, doorbuttons were made out of nuts), then ringing a doorbell would no longer mean what it now does. But as things now stand, we can say that the bell would not be ringing unless someone was at the door. It therefore indicates that someone is at the door. But this subjunctively expressed dependency between ringing the bell and someone's presence at the door, though not a coincidence, is not grounded in natural law either." (Dretske 1988, p. 57)

So here we have a case of indication but not in virtue of some natural law. Dretske insists that mere (accidental) correlation is insufficient for indication, however:

> "There must be some condition, lawful or otherwise that explains the persistence of the correlation." (Dretske 1988, p. 57)

To this, Millikan counters:

> "But, of course, if the condition that explains the correlation is not lawful but 'otherwise', then it is grounded in mere facts about the state conditions characterising the world at certain

> times and places — either conditions at the source or existent channel conditions. It has the status merely of a local statistical frequency — based lawfully, perhaps, hence explainably upon prior local statistical frequencies, but that does not change its essential nature as a statistical frequency." (Millikan 2000, Appx. B, p. 222)

So we have two senses in which statistical frequencies might enter into a notion of information, even one that is putatively given as being tied to conditional (nomic) probabilities of 1. Here is an example of Millikan's (originally from Grice): 'Those spots mean measles' might be read as saying that those kinds of spots indicate (naturally signify) that the person with those spots has measles. First, however, that we are in a set of (channel) conditions in which this holds may only be a matter of statistical frequency. Given other (channel) conditions, perhaps those spots would mean that some other virus is present. Second, it may only be a matter of chance that we happen to be in a situation in which only measles cause those kinds of spots. Were we in another part of the world where both measles and some other virus cause the same kinds of spots, then those spots would not mean (indicate) measles.

These considerations alone suggest that Dretske's later position allows for an alternative, non-nomic, conception of information. Further arguments for this conception of information will be provided below.

4.1.2 Types of Representation

Dretske (1988) classifies signs into different categories of representation. He calls these *Type I*, *Type II* and *Type III representations*. He places natural language signs ("to some degree") into the category of Type I representations (Dretske 1988, p. 53). Type I representations represent only insofar as we assign them functions. They do not carry any information intrinsically. That Dretske places language (to some degree) here is not surprising. How could what linguistic signs represent be governed by natural laws?

However, when we look at Dretske's example of Type I representation, and also try to make sense of linguistic information, there are problems. His example is a case of assigning coins and pieces of popcorn the role of representing basketball players in a game he has watched. Symbols are "either explicitly or implicitly, assigned indicator functions, functions they have no intrinsic power to perform. We give them their functions, and we (when it suits our purposes) see to it that they are used in accordance with this function." (Dretske 1988, p. 53). This raises a puzzle. What could Dretske mean by 'we'? Should we view the 'we' as somehow of one mind?

Maybe what Dretske has in mind here could be captured by something like Putnam's division of labour (Putnam 1975). Designers of devices determine what they are for and WE defer to them. However, this does not seem right either. For everyday terms ('green', 'tall' etc.) we do not seem to defer to any experts (nor was this the idea in Putnam). Experts may well give functions to some technical terms, but for words which need no expertise beyond competence in a language, could it be that we ASSIGN words a function? Certainly not explicitly. Dictionaries may, to some extent, describe what the functions of words are, but dictionaries do not assign these functions. Even if they did, Dretske can't just mean *lexicographers* by 'we'.

Yet, it isn't easy to grasp what it might be for some group of people to IMPLICITLY assign a function to a word. Words might develop reasonably stable functions through repeated uses by large numbers of speakers (for instance, that 'green' is, roughly, used to describe some shades but not others), but that does not mean that we (the speakers) assign the words these functions.[3]

Some views of utterances appeal to intentions of speakers (for example (Grice 1957)). Yet such views are not meant to account for the information that words encode. Some aspects of what words indicate is beyond the power of what any individual can intend. I can intend all I want that 'Polar bears live in the Arctic' indicates *Squirrels live in forests*, but (unless circumstances are very special) using those words will not mean that.

For natural language signs, the 'we' doing the assigning is thoroughly opaque and before we accept this view of indicating, it must be clarified further. However, even if we accepted a clarified version of the account, it is not clear that linguistic signs must represent information in a different way to natural signs. It is only Dretske's account of natural information that leads him to conclude that Type I representations do not always convey information. In §4.2, I will argue that his account of natural information is not warranted.

Dretske's Type I representations contrast with his *Type II representations*. Type II signs carry natural information (think of the tree rings indicating

[3]By the end of this chapter, one way to understand 'implicit assingment' will be elaborated. However, on this understanding, words carry the same kind of information as natural signs.

more than just the age of the tree), but, for these kinds of signs, we might not be prepared to recognise everything that is indicated as what the sign means: "A sign is given the job of doing what it (suitably deployed) can already do" (Dretske 1988, p. 59).[4] In light of the problems with viewing linguistic signs as Type I representations, I propose that we allow linguistic signs to carry information just like any other sign. As noted in the last chapter, not all information carried need be seen as part of the meaning of that sign. However, some of that information is what linguistic signs are used to represent. In other words, I propose that we treat natural language signs as Type II representations. The basic idea is that, once entrenched in uses, written down and practiced, natural language signs are in fact indicators much like any other (natural) signs. We may not wish to accept all of what linguistic signs indicate as part of what those signs mean. However, if some part of what they indicate is equivalent to what they mean, then we might be some way towards a naturalistically acceptable view of language on which agents use some of what words indicate routinely enough for those signs to have that as their meaning.

To get to this conclusion, however, will take some work. I shall argue, along with Millikan, that, for natural signs, we should not adopt Dretske's account of indicating based on natural laws, but instead, an account based on environmental correlations. That will pave the way for motivating a softer notion of information to be carried by natural language signs too.

4.2 Against the hard conception

On the basis of the distinctions made in §4.1.1 (information based on certainties, frequencies affecting the channel, and frequencies affecting the source), Millikan gives a series of arguments for why we should not expect certainties to be carried by natural signs (so as to be utilised by organisms in their intentional representations). I will look at two of these arguments, which I will extend to apply to communication and language. Finally, in §4.3, I will argue that, in the case of language, we should only expect information to be grounded in statistical frequencies.

[4] Whereas Type III representations are equivalent to what I refer to as 'natural signs'.

4.2.1 Situatedness

To account for Type II misrepresentation, Dretske appeals to teleology. He gives an example of an altimeter. An altimeter carries a lot of information such as the air pressure in its locale and its altitude above sea level. Even if the dial on the altimeter is not set right, the device still carries this information since it still indicates, say, altitude, even if one cannot read this height from the dial. The device still carries this information. If we knew how much the dial was out by (how it was callibrated), we could still decode the device to learn the altitude. In this sense, the altimeter cannot mis-indicate anything. Of all the information they carry, Dretske tells us, altimeters are assigned, by us, the function of representing altitude above sea level. The representational character of such devices is "interest-relative and purpose dependent" (Dretske 1988, p. 67). It is via this assigned function that Dretske sees a possibility of misrepresentation:

> "[T]he only way a system of natural signs can misrepresent anything is if the signs that serve as its representational elements fail to indicate something that they are SUPPOSED to indicate. And what they are SUPPOSED to indicate is what WE, for purposes of our own, and independent of a sign's success in carrying out its mission on particular occasions, REGARD them as having (or give them) the job of doing. Without US there are no standards for measuring failure, nothing the system fails to do that it is supposed to do. Although the actual failures aren't our failures, the standards (functions) that make them failures are our standards." (Dretske 1988, pp. 66-7) (Dretske's emphasis)

Millikan, albeit tentatively, suggests that the picture that Dretske has in mind in the above quote is one whereby misrepresentation can only occur when systems fail (malfunction) (Millikan 2000, p. 230-1). This view cannot, Millikan claims, be right. Misrepresentation is not always malfunction, but can be generated by an uncooperative environment. Our representational systems can function perfectly normally, but if their environment is altered, their normal functioning may begin to lead us astray. In Dretske (1988) however, he seems to have a more nuanced view than the one Millikan assigns him. Dretske gives two ways in which misrepresentation can occur. One of these is malfunction: "a device can fail to do what it was designed to do" or devices can "wear out and no longer retain the power to indicate what it is their function to indicate" (Dretske 1988, p. 67). However, Dretske also allows for just the kind of uncooperative environmental failure that Millikan alludes to. In Dretske's words, devices can be "used in circumstances that curtail their ability to indicate what they were designed to indicate" (Dretske 1988, p. 67).

Here we can use a point of Millikan's to bring out the problem with Dretske's conception. Environments may be "rife with decoy channels" (Millikan 2000, p. 231). If carrying information is the carrying of nomic certainty, as Dretske would have it, then the chances of anything carrying information are slim. Recall what Dretske said about doorbells and squirrels. That there are or may be decoy channels breaks the certainty needed for information carrying to occur. If the perfect *situatedness* of signs (in channel conditions and source situations where no decoy channels occur) is a necessary condition for a sign to carry information, then rarely, if ever, will signs carry information.

One might think that words cannot carry natural information simply because it makes little sense to think of this information as nomically underpinned. One might also come to this conclusion because words can be used to misrepresent how things are, and so there can be no nomic certainty tying signal to signified. We might, nonetheless, engage in a little fantasy. Suppose that people are such that in the presence of certain conditions external to themselves, they always respond by uttering certain words. Imagine that this connection is, in some sense, underpinned by natural law. This fantasy effectively rules out misrepresentation. However, this is exactly where Millikan's point about decoy channels comes to bear. In such a world, the information carrying capabilities of a signal would almost always be undermined by the presence of improbable although possible decoy channels, and they would almost always be undermined by statistically improbable, although possible, elements of a source situation. This would mean that, even in the good case where there is no misrepresentation, linguistic signs would rarely, if ever, carry information and so rarely, if ever, would linguistic signs do what they are meant to do.

4.2.2 Insufficiency

Information based on natural laws, Millikan argues, does not tell creatures everything they need to know to survive:

> "The mouse, for example, needs to know when there is a hawk over head, but there are no natural laws that apply to hawks and

> hawks only... It is the properties of objects like hawks and cows that enter into natural laws" (Millikan 2000, p. 227)

Two points are being made here. (i) That individuals do not feature as the *relata* of natural laws. (ii) Species don't either. Laws govern properties, and it is only in virtue of having properties that individuals and species can be governed by laws. However, species cannot be defined in terms of their necessary and sufficient properties,[5] and it seems dubious that individuals can be either. But organisms need to use information about other individuals and other species to survive. Therefore, laws seem to be the wrong thing to ground the information that organisms need to survive in their environments.

This point can be extended to communicative signals. I will take two cases: Communicative rabbit thumps, and human verbal communication.

Rabbit thumping behaviour has emerged via an evolutionary process as a way for rabbits to improve their chances of survival. Under certain conditions, one rabbit thumps. On hearing these thumps, other rabbits engage in fox avoidance behaviour. Sometimes a rabbit thumps when there is a proximal fox, and other times they thump despite there being no proximal fox. This means that sometimes, what causes a rabbit thump isn't always a fox. Let us consider, similarly to the mouse case, whether other rabbits need information other than the nomic variety. Suppose for a moment, though, that some rabbits only thump when they encounter a sure sign of a fox (that only foxes cause thumps). Given that these *fox-perfect* rabbits do not have super-leporid senses, fox-perfect rabbits would sometimes not thump when normal rabbits do. For example, if a flash of reddish brown fur (often/sometimes) causes normal rabbits to thump, these other rabbits would not. This is because, in rabbit environments, although reddish-brown fur may be correlated with foxes, it is also correlated with other things such as deer. The fox-perfect rabbits would have to be exposed to something that did not have just fox correlation, but also had no other correlation (deer correlation etc.).

We should expect fox-perfect rabbits to produce no false positives, whereas normal rabbits produce some. More worryingly though, fox-perfect rabbits would produce more false negatives. To use a sensor metaphor, the fox-perfect rabbits have a less sensitive sensor. The actuality of this will be that foxes may get closer without detection. More false negatives means

[5] See (Millikan 1984). Millikan argues that no biological categories can be defined in terms of necessary and sufficient conditions.

more rabbits in the community get eaten. With more sensitive sensors in their community, those rabbits might otherwise have had the chance to engage in fox avoidance behaviour.

For the information decoded by a non-malfunctioning fox detector to be grounded in nomic certainty, rabbits would have to be fox-perfect rabbits, but they are not fox-perfect rabbits. If we allow that thumps carry any information at all, this means that the information that rabbit thumps carry is not nomically certain information. The point here is that we do not and should not need to view all cases of misrepresenting as cases of malfunctioning. This relates to discussion earlier on how even natural signs can go wrong. Sometimes, a natural sign can mis-indicate because it is situated in the wrong conditions. For the rabbit case, the point is that, even if conditions are right, the information carried can't be nomic.[6]

One way to put this is that normal rabbit thumps carry information based not on nomic certainty, but on correlation. This correlation is grounded in certain facts about rabbits and the environments they inhabit. Rabbit avoidance behaviour is therefore also grounded in these correlations. This sensing, thumping, and reacting behaviour is a form of coordinated behaviour within rabbit communities that has been developed to aid the survival of the rabbit population. In the sense that this is communication, on the information theoretic model of communication, the rabbits are, by thumping, transmitting information. Clearly, hearing a thump does not provide a rabbit with the information that a fox is in the area, but grounded in the kinds of correlations found within rabbit environments, since it seems to be information of some kind, a plausible characterisation of it would be the information that a fox might be/probably is in the area.

Of course, on that view of information, rabbit thumps transmit more information than this.[7] This picture of thumps makes them now look just like Type II signs. These signs carry a wealth of natural information. Their meaning is part of this information, namely, the part that rabbits use: that a fox is present.

The basic idea can be carried over into human communication. We start from the idea that communication involves, at least in part, the transmitting of information. Dretske's approach (albeit not applied to language) is to posit a strong notion of information and come up with an explanation of misrepresentation (the sign is not situated properly, or the system is

[6] Cf. The argument in (Millikan 2004, pp. 33-4).

[7] Amongst other things, that a fox might not be in the area, or that a deer might be in the area.

malfunctioning in some way). However, nothing forces us down this path, and, in fact, there seem to be reasons to not go down it (that kind of information is seldom available, and is less than we need to get by). Instead we can simply call information whatever it is that gets communicated, and then try to say something about THAT. On this view, information cannot deal in certainties, but it is not random either. The sentence, 'Lucy is out' can be used by her housemate in reponse to a request for information about where Lucy is, but rarely if ever will this give us certainty about Lucy's position, nomic or otherwise. Nonetheless, the signal is informative, and the source (Lucy's housemate) is frequently reliable. Receiving the signal gives us excellent reason to believe, on the basis of past success in our communicative enterprises, that Lucy is not at home even if it turns out, this time, not to be the case.

However, there is a significant difference between the uncertainty found in rabbit communication and human communication. In the case of rabbits, the uncertainty left by the thumping signal relates to whether or not a fox is present. This kind of uncertainty is also present in the Lucy example (whether or not she is at home). However, there is another form of uncertainty in human communication. Given the signal 'Lucy is out', there may be any of many things Lucy is doing. Nonetheless, given the correlations that pertain between uses of expressions like 'out' and the kinds of things people do when they are out, some of these things may be more probable than others. For example, if someone is described as being out, then it is, on the basis of how this expression is used, more probable that they are not too distant and going to return in the near future than, say, away somewhere distant for the next week, fortnight or month. This difference in uncertainty will be elaborated on in later chapters. The former, I shall refer to as *worldly uncertainty*, the later as *metalinguistic uncertainty*.

4.3 Correlations Persist for Good Reasons

Weaker than certainty is uncertainty. Just as naturalistically acceptable as nomic certainty is a signal carrying information in virtue of being grounded in statistical correlations. That said, we do not want any old accidental correlations to count as information. These correlations need to persist for a good reason. Finally, we will look at one account of exactly why correlations between (uses of) the word and the world persist.

In the appendix to her (2000), Millikan is primarily concerned with criticising Dretske's account of natural information based on nomic certainties (what she calls 'informationL') and on promoting an alternative, softer, notion based on statistical frequencies (what she calls informationC). Our concern is not merely natural information, but also intentional, linguistic signs and the information that they carry. First I will outline Millikan's conception of informationC, then I will suggest how it can be applied to linguistic signs.

For natural signs, Millikan recognises the importance of natural laws in the account. Her crucial insight, however, is that a sign can be related to some state in the world partly because the sign is situated in a law governed environment, even though this does not make the connection between sign and signified in itself law governed. This long passage makes the point with admirable clarity.

> "If the frequency of the black balls in the urn today is 1, and if nothing disturbs the urn, then by natural necessity it follows that the frequency of balls in the urn tomorrow is 1. That does not change the probability of being black if a ball in the urn into a probability of some kind other than statistical frequency. It does not help being-a-ball-in-the-urn to carry the informationL being black.
>
> But it does do something else. It explains how, by sampling the urn today and adjusting expectations of colour accordingly, this adjustment in expectation can turn out to be adequate to my experience tomorrow, NOT BY ACCIDENT BUT FOR GOOD REASON. Many statistical frequencies persist over time in accordance with natural necessity, and many produce correlate statistical frequencies among causally related things, in accordance with natural necessity. If measles are producing spots like that in this community today, then measles will PROBABLY be producing spots like that in this community tomorrow. Measles, after all, are contagious. And if a nose like that is correlated with the presence of Johnny today it will probably be correlated with the presence of Johnny tomorrow. Johnny's nose, after all, tends to sustain both its shape and attachment to Johnny. There are no laws that concern individuals as such, but there are many correlations that do. Notice, however, that whether the persistence of a correlation may be explained in this sort of way DOES NOT DEPEND ON ITS BEING A PERFECT CORRELATION. Conditional probabilities of 1

> have nothing to do with the matter." (Millikan 2000, pp 234-5)
> My emphasis.

This background of a law-governed world is also where intentional linguistic signs are situated. But linguistic signs do not seem to be supported by these laws in the same way as non-intentional natural signs. If utterances of the word 'green' are correlated with the presence of/or with reference to objects of some shade(s) today, then whilst it is no doubt the case that productions of the sign will be correlated similarly tomorrow, this does not seem to be a persistence of the object through time (like the black balls in the urn or Johnny's nose). Nor does it seem to be because of causal processes (like the measles). What we need to fill out the linguistic picture is a notion of why correlations between (uses of) linguistic signs and the kinds of things they signify persist, not accidentally, but for good reason.

Natural language has already been described as having communication as one of its central functions (more accurately, it allows users of the language to communicate). Communication was analysed, in part, as the exchange of information. Missing from this picture is any account of why we exchange information (why we communicate).

We do not need to commit to any view about the cooperative (as opposed to the merely coordinative) nature of communication (see ch 3) in order to talk about the goals, aims and purposes of certain communicative actions and endeavours. Words and language help us to do things that we would otherwise be unable to do. To name just a few: To find out information not directly available to us, to provide others with information that they lack, to have actions performed that we are either not able or not willing to do alone, and to partake in sophisticated joint endeavours. Instances of these goals and actions are reasons for engaging in linguistic activity if linguistic activity will help us to achieve them. It is simply this that will form the basis for the good reasons that correlations between linguistic signs and what they signify persist. To help fill out this picture, I will again turn to Ruth Millikan and her account of natural conventions (Millikan 2005a).

Millikan offers an account of conventions that is much simpler than the kind of view found in (Lewis 1969). Millikan defines conventions as behaviour that is reproduced due to a weight of precedent, not because of (intrinsic) function. Reproduction due to weight of precedent simply indicates that a reason for an individual reproducing a sign can simply be that that sign has been used for some effect before, and that the agent wishes to bring about that effect. Behaviour is witnessed and reproduced such that were the witnessed behaviour different, the reproduced behaviour would be too. Millikan should not be read as making a strong modal claim here. Implicit in her examples is that there should be some sort of connection between an action and a reproduction of that action. It is, of course, possible that had the reproduced action stayed the same, but had conditions differed slightly, the reproduction would have changed. Or, in cases where the reproduction is anyway slightly different from the reproduced action, it is possible that, had the reproduced action differed, the reproduction would not have changed.

Reproductions of actions should not, if we are to call them conventional, be due to the function of the action being reproduced:

> "Handing down a skill is not, as such, proliferating a convention. I learned from my mother, and she from hers, to open a stuck jar lid, by first immersing it in hot water. Opening jars in this way is not thereby 'conventional'. To be thought of as conventional, a reproduced pattern must be perceived as proliferated due, in important part, to weight of precedent, not to its intrinsically superior capacity to produce a desired result, or due, say, to ignorance of any alternatives" (Millikan 2005a, p. 7)

Here we have the beginnings of a picture on which what certain signs, linguistic or otherwise, signify can be conventional. Importantly, however, 'conventional' is here being used merely as implying the reproduction of behaviour which proliferates due to a weight of precedent. Firstly, this does not mean that patterns are not reproduced for any reason. They can be reproduced simply because they have been used to achieve things in the past. Secondly, the link between pattern and effect may have to be learnt (more on this in chapter 5).

A final feature of Millikan's picture will allow us to account for the kind of propagation of a correlation between linguistic sign and signified state of affairs that we needed in the last section. Conventions, linguistic or otherwise, do not need to be exceptionless (Millikan 2005a, p. 14 & p. 16). That someone uses a sentence or some words on some occasion does not make it so that what the words signify is the case or even is almost always/very probably the case. All that matters is that what those words conventionally signify is observed regularly enough for the correlation to have precedent, to be propagated, and to be entrenched (gain a greater weight of precedent) via reproduction.

With the three pieces in place (reproduction, weight of precedent and non-exceptionlessness), we now have a way of interpreting Dretske's intuitions about language being classified as a Type I representation (one that represents only in virtue of assigned function).

It must be right, in line with Dretske, that linguistic signs do not signify what they do because of some intrinsic function or characteristic. However, in light of the statistical notion of information, the intrinsic/extrinsic distinction is perhaps not the best one to use. On the statistical account, signs (linguistic or otherwise) can carry information if there is a non-accidental correlation between sign and signified. Millikan's account of conventions provides the explanation for why, with linguistic signs, these correlations are not accidental. They are propagated for a reason. However, nothing intrinsic to the sign plays a role. Indeed, if what the sign signified were entirely due to some intrinsic feature, it's reproduction would not be conventional.

Contra Dretske, we needn't go so far as to think that linguistic signs are assigned their functions by 'us'. Signs may have been used in the past to successfully bring about certain results. Merely in virtue of learning that certain patterns or signs can bring about those results, albeit without any certainty involved, is enough to propagate a reproduction of those signs and patterns. This is compatible with the functions of signs not being assigned. For example, uses of novel signs can still bring about a desired effect. In turn, this may lead others to reproduce the sign to bring about similar outcomes. Simply by being used to do a particular thing is enough for a correlation to build, and so for the sign to get a (conventional) meaning. In this sense, meaning can emerge as the result of interactions between agents and reproductions of those actions.[8]

4.4 Summary

On the account proposed, all signs carry information and this information is grounded in statistical correlations, not in nomic certainties. However, what separates natural signs from linguistic signs are the reasons for why the statistical correlations which characterise the information they carry persist. In the natural case, the reason why the position of Johnny's nose continues in correlating with the position of the rest of Johnny's face will probably involve some natural laws which explain the way molecules are joined

[8] It is possible that this approximate picture (functions emerging over reproduced uses) is what Dretske means by the claim that what a sign indicates can be implicitly assigned. If so, my quibble may be reducible to a dispute about terminology.

together and the forces needed to break them asunder (as well as statistical probabilities relating to the frequency with which such forces are applied to people's noses). The correlations that characterise the information carried by linguistic signs persist for different reasons. These reasons are based on the way that human beings repeat and reuse signs that are successful in bringing about particular ends.

The details of how to characterise semantic information and represent it formally have yet to be given (this will feature in chapter 6). As an informal characterisation, the picture will look as follows. When used to describe some object, calling that object, say, 'green' will tend to correlate more strongly with it being some shades rather than others. Information reflects statistical correlation. So the information one can decode from such a description will be that the object is PROBABLY one or other of some shades, and probably not any of some others. This picture does not (yet) feature truth. If things are some way that was improbable given the signal, then we may well want to say that what was said is not true. However, nothing in the words themselves or the information they carry answers that question for us. If the information that linguistic signs carry reflects merely what they tend to be used to describe or indicate, then the signs themselves (the language itself) can tell no lies (nor any truths). It is people that speak falsely or truly by using words, not the words themselves that say something true or false. Just how truth features on a soft notion of information (and what kind of logic we should use) will be the central topic of chapters 7 and 8.

5

Semantic Learning and Information

We can now turn to the other central criterion for developing a semantics for vague terms: learning and learnability. Whereas most of the last two chapters were concerned with communication (what is communicated, and what the nature of what is communicated should be taken to be), now we will consider what the meanings of terms (or the information they carry) must be like if they, the meanings, are to be learnable. Again, just as in chapter 2, we shall see that this consideration militates against a view of meaning and information as truth-bearing. It will also support a conception of meaning based on uncertainty and the inference of probabilistic generalisations.

In §5.1, I will introduce some work done on learning models and the main problem that they try to solve: the learning bottleneck. These models assume a basic mapping between strings and meanings, and so they are not focussed on learning meanings as such. In §5.2, I will suggest a way for the insights of these models to be applied to semantic learning. Doing so will provide some constraints on what the meanings of terms must be like to be learnable. Finally, in §§5.3-5.4, the conclusions of the current and the previous two chapters will be discussed.

5.1 Iterated Learning Models and Bottlenecks

In recent years, a method of modelling how languages develop over generations of learners has been developed (see for example (Brighton and Kirby 2001), (Kirby and Hurford 2002) and (Kirby 2007)). These *iterated learning models* (ILMs) were designed to address a fundamental challenge in the field of language learning: the *learning bottleneck*. Learning bottlenecks can relate to a variety of domains. Simply put, a bottleneck problem arises in any form of learning where learners are not exposed to everything they need to learn in the learning period itself. That means that something may be required of a learner after their learning period, which they have no direct experience of.

In a simple ILM, languages are modelled as collections of string-meaning pairs.[1] The strings are simply the possible strings of a language (i.e. sentences) and each string expresses some meaning. The bottleneck arises because learners of a language are not exposed to every string-meaning pair. After the learning period, learners become competent speakers, but have not experienced how to express every meaning that the language is capable of expressing.

I will summarise just one of the results, presented in (Kirby and Hurford 2002), to come out of ILM research. Kirby and Hurford model language change in a simplified format whereby, at any generation, there is a single learner and a single competent speaker. During the learning phase, language learners are exposed to a finite number of string-meaning pairs.[2] This exposure is meant to model the sample of a competent speaker's language that a learner has as primary linguistic data in their learning phase. It is built into the model that learners are efficiently able to learn string-meaning pairs simply by being exposed to instances of strings during the training phase. Put another way, just by being exposed to a string/sentence, a learner learns the sentence and its meaning. By the end of the training phase, the learner becomes a competent speaker, and a new learner is introduced that is exposed to a sample of the old learner's (the new competent speaker's) language. The goal for each new learner is to be able to produce a string for any meaning that may be demanded. Meanings may not be ones that the learner has encountered in their learning phase, so they may have to produce strings that they have not previously been exposed to. It is in

[1] In the ILM literature these are referred to as 'form-meaning pairs'.

[2] Informally, this means that learners are given a package of a sentence of the target language plus some kind of meaning representation of that sentence (such as a proposition written in first-order logic).

this sense that these models are mainly syntactic, not semantic. Meanings are assumed from the outset as fully defined in some formal language. Learners efficiently learn meanings simply by being exposed to strings with that meaning. But the learning task they are engaged in is one wherein they must learn how strings should be structured to express meanings. Learners must assign parts of strings to parts of meanings. For example, if $W(a)$ is the meaning of 'Rory walks' and $W(b)$ is the meaning of 'Sarah walks', learners must identify which terms ('walks', 'Rory', 'Sarah') refer to W, a, and b. In §5.2, we will consider a learning task where no presupposed class of meanings (what, in the literature is called a *meaning space*) is assumed. There, the learning task will be semantic. Given some multiple exposures to a term, learners will have to infer a meaning for that term.

5.1.1 Success Criteria for ILMs

In ILMs, over many generations of speakers and learners, a model is judged to be successful if the language that emerges over these generations bears two main characteristics[3]: *Expressiveness* and *Stability*. A language is expressive if competent speakers are able to provide a string for any meaning that is in the meaning space.[4] A language is stable if each learner ends up with a reasonably similar set of string-meaning pairs as the competent speaker from whose language their training sample is provided. The decision not to require complete stability is motivated by the fact that NLs do change in small ways over generations of learners.

A challenge arises for both of these success criteria: the learning bottleneck. The learning bottleneck means a learner may not be acquainted with every string-meaning pair in their learning phase. If a learner is not acquainted with a way to express a certain meaning after completing their learning phase, then their language will not be fully expressive. Expressibility can be accommodated by allowing speakers a degree of innovation. If they have not encountered a way to express a certain meaning, they are allowed to use a novel form that hasn't been learnt to express it. However, this solution creates stability problems since it is highly likely that innovative language use will not match the relevant string-meaning pair of the elder generation's language.

The solution that Kirby and Hurford propose is to endow learners with

[3]There are, in fact, more than two such characteristics in more advanced models, but I shall stick with only these two here.

[4]Expressiveness does not require that the string be correct/appropriate for the meaning.

more sophisticated learning abilities. As well as learning string-meaning pairs (storing facts), they are also able to draw generalisations over how parts of forms align with parts of meanings.

Learners therefore have two learning strategies. (i) To simply store string-meaning pairs. (ii) To form generalisations over how parts of strings align with parts of meanings. The result of learners employing these two learning strategies is that, over generations, a compositional and recursive language emerges. Put another way, languages develop a syntax.[5] This begins to solve the problems associated with the bottleneck, since the generalisations that learners draw are often, in fact, over-generalisations that allow them to express meanings that they were not exposed to in the learning phase. After many generations, whilst the facts learnt may differ, provided that the generalisations drawn by learners converge, language remains reasonably stable, since the language of each generation will be characterisable by the same string-meaning pairs.

5.2 The Semantic Bottleneck

The argument in this section will follow from Kirby *et al*'s work by analogy. Instead of focussing on a whole language, one can think of a similar process going on for particular terms, predicates, say. In this analogous process, a user of the predicate will, during their training phase, experience its use in a variety of contexts. These contexts will be of two types: (1) The predicate will be used to apply to different objects. (2) The predicate will be used to apply to the same object in different external conditions.

In each context, the user learns a predicate-meaning pair. In fact, more accurately, they learn a predicate-object pair, since learners should not be understood as learning a full-blown meaning for the predicate with every use that they are exposed to. What is being learnt is, rather, that some object falls under that predicate. The object may itself be classified in some way. For example, if 'red' is the classifier being learnt, then one may learn that some object, a postbox, is classified using 'red'. It is therefore being assumed that a learner can identify some object or objects that are being described, and store which terms are being used to classify them. For example, the learner can identify an object as a postbox and store that it has been classified as red.

[5]Here syntax is being understood in the Chomskian paradigm as the set of generalisations (rules) learnt. Nothing in my argument to follow turns on this notion of syntax being adopted.

In parallel to Kirby *et al.*, we can define some success criteria. A predicate user will be successful with her own use of the predicate after the training phase if she satisfies two requirements. *Expressiveness:* If, when faced with the requirement of classifying some object in an external context, she is able to select a predicate to apply to that object (i.e. can draw on her learning phase data in order to make a classification). *Stability:* Three criteria of success emerge for stability. (i) When faced with a task of classifying some previously classified object in a previously unencountered external context, the speaker will, on the whole, use the same predicate as in similar contexts that have been experienced. (ii) When faced with classifying a previously unclassified object (or type of object) in a previously experienced type of context, the speaker will, on the whole, use the same predicate as with similar objects previously classified. (iii) Different learners will attain reasonably similar behaviour when given a different sample of training data. The first two principles of stability are reminiscent of the principle of tolerance found in (Wright 1975): If minor changes in objects or contexts arise, predicate use should not, on the whole, vary too greatly. Another way of describing this is that predicates (or their application) should be optimised to survive over minor shifts in context.

A learning bottleneck arises here, too. Since, in their learning phase, a learner will only be exposed to a limited number of contexts and predicate-object pairs, they may not have a way of describing either an unencountered object in a context (previously encountered or not), or a previously encountered object in an unencountered context. This creates a challenge for expressiveness. Analogous to the allowance of innovation found in Kirby and Hurford's model, we might allow the language user to be innovative. However, if their selection is effectively random, then a problem is created for stability.

Just as in the previous case, I propose that a solution to this problem lies in endowing learners with inductive learning abilities. We must assume that, not only will learners be able to store predicate-object pairs, but they will be able to infer generalisations about which predicates are used to classify which objects/events in various contexts. Whereas generalisations for Kirby and Hurford took the form of syntactic generative rules, the generalisations here will be different. Given the nature of the task for the speaker (to express certain meanings/apply a predicate to an object in a context), they will be able to form at least two different kinds of generalisations. The first will concern application to objects. The more similar an observed object is to one previously experienced, the greater the weight the speaker should assign to the same predicates applying to it.[6] The second kind of generalisation will concern objects in an external context: The more similar a context is to one previously experienced, the more weight the speaker should attribute to similar objects in that context being classified the same way as in the training phase.

Just as the generalisations made by learners in Kirby and Hurford's model could be taken to represent the syntax of their language[7], the generalisations made by learners of predicates about their application can be taken to represent the meanings/the semantics of those predicates.

How are these generalisations best represented? One option would be to represent them as declarative rules. For example, learners might form a generalisation such as, *apply predicate F to object o if o is similar to o' in respect y to degree x or more, where o' has been experienced as being classified as F.* This kind of generalisation is problematic: (a) The rule would have the effect of creating a sharp cut off point for every predicate a set distance (x) from the least central case that the learner has experienced. However, the cut-off points between learners would vary depending on their training data. This amounts to a type (iii) stability problem. (b) The rule is too simple. A learner would also need to be able to track the effect of context by encoding how some feature of the context affects which objects can be classified by a predicate. Any kind of similarity relations between contexts would themselves require more declarative rules, each one with another wholly arbitrary cut off. (c) It is far from clear how the rules could themselves be learnt. A learner would need to identify which y, and which value for x, should be identified as relevant for each F.

(a) means that different learners would be unlikely to learn the same representations. *A fortiori*, if there were some TRUE meaning for the classifier, learners would be unlikely to learn it. This sort of problem is one that we encountered already at the end of chapter 2: When idealisations are made about there being, relative to some context, a simple answer as to whether it is true/false that some classifier applies to an object, and, where this question is held to be answered by the meaning of the classifier itself, it is unlikely that the entire meaning of the classifier will be learnt. This is because data one gets as a learner will not be sufficient to allow the learning of the exact truth conditions of the classifier. (b) and (c) indicate that even

[6]It is being assumed that learners have the ability to judge how similar objects are.

[7]Again, with the same hedged remark as above with respect to how syntax should be represented.

single learners face intractable learning problems if the generalisations learnt are of the above form.

For effective classification, a meaning representation is needed that incorporates the weighing of different factors, and that can reflect degrees of similarity between objects and contexts. What is therefore needed is a meaning representation that reflects how much/how strong a reason an agent has for classifying an object one way or another.

Predicate learning tasks involve identifying relevant features of contexts and objects from very messy data. Furthermore, generalisations relating to applying predicates to objects must be inductively learnt from a limited number of cases. This kind of learning task cries out to be modelled probabilistically.[8] Probabilistic models can reflect the weight of different factors, such as properties of contexts and objects. Furthermore, work on Bayesian learning has provided some promising suggestions for how appropriate structuring of data can be inductively learnt (Kemp and Tenenbaum 2008).

5.3 The Proposal in Brief

A full account of how I propose to model the semantics of vague classifiers will be given in the next chapter. Here, I will briefly discuss how the above discussion on learning, and our previous discussion on communication, will influence the picture I will defend.

The conclusion from the previous section was that if classifier learning is modelled probabilistically, then meaning representations should also be probabilistic. But probabilities over what? The output of this learning process will involve the properties that objects have.[9] For example, the result of learning a classifier such as 'green' will involve generalising over the shades of colour that learners observe objects (which are classified with 'green') to have. This is because shade of colour will be the relevant property identified over multiple instances of 'green' classification.[10] Properties that

[8]No claims are being made about whether actual learners have probabilistic mechanisms in their biological make-up. See (Tenenbaum et al. 2011) and (Griffiths et al. 2008) for discussion and an introduction to Bayesian learning research in cognitive science.

[9]In one recent proposal for (nominal) classifier learning (Cooper et al. 2013), the classifier 'apple' is learned via inputs of colour and shape. However, for learning colour terms, one might have to use more determinable or fine grained properties such as shades of objects. This is because, if the topic of analysis is how one learns colour terms, and since colour terms exhibit the kind of flexibility characteristic of many natural language expressions (chapter 2), one should not assume that a learner can automatically identify what is and what is not red. After all, part of what one must learn is what shades of things it is appropriate to call, say, 'red', in which situations.

[10]This is an empirical claim, but I take it to be a fairly safe assumption.

have been more frequently experienced as being classified as green will get higher values of expectation when ‘green’ is subsequently used.[11] The representation of this will be a probability distribution. For example, the interpretation of ‘Mike’s shorts are green’ will be, in part, a probability distribution over the shades Mike’s shorts could be, weighted appropriately to the way ‘green’ is used in English.

Probability distributions formally represent just the kind of natural information that was argued for in chapters 3 and 4. There, I held that natural information is grounded in statistical frequency and that, so seen, human linguistic behaviour can be seen as carrying natural information. The reproduced conventional correlations (chapter 4) that hold between uses of words and types of situations are, I suggest, what reflect the meanings of our words. From a communication standpoint, what counts as meaning is restricted by our purposes and enforced via reproduction (conventionalised). From a learning point of view, these statistical correlations can be learnt and estimated by applying statistical learning strategies to data provided by experiences of repeated uses of terms.

Of course, further information provided either lexically or from the wider context will affect the distributions that language users have. For example, if shorts tend to be some shades of green rather than others, the interpretation of the above sentence will reflect this information (if it has been learnt in the leaning of the nominal classifier ‘shorts’). Alternatively, if one knows that Mike has been rolling about in the grass, this may also effect one’s expectations as to the shade his shorts are (in part, this information will come from having learnt the term ‘grass’). How this kind of information can be modelled as interacting will be elaborated upon in the next two chapters.

5.4 Summary: Communication, Learning and Vagueness

Language users face a bottleneck problem not only in a syntactic sense, but also in a semantic one. New contexts and previously unencountered objects pose a challenge for a language user in the way that they classify the world. Probabilistic learning seems to be an attractive way of addressing this challenge.

However, probabilistic representations of meaning leave speakers and

[11]This is slightly inaccurate. In fact, objects get classified and the learning process will associate common properties with classifier meaning.

hearers with a lot of uncertainty. For example, the world might be very many different ways given that the speaker has classified o as F. Furthermore, given some observed way for the world to be, a speaker might be uncertain whether or not to use F to describe it. A potential problem is created by this conclusion. I have, in effect, argued that learning how to apply predicates creates uncertainty over whether to apply a predicate to an object in a context. This uncertainty propagates through to a hearer, who is thereby left with uncertainty over how the world is, given that it has been described a certain way. This latter uncertainty could be seen to create a resolution problem for the hearer, since they will not know how the world is according to the speaker. We will need to add further details to the picture to address this resolution problem.

In chapter 7, I will give an outline of how context and purposes of speakers and hearers can help to constrain the kinds of uncertainty we are faced with. A possible account of how to model context-sensitivity becomes available by taking an information theoretic approach. If the kind of information that one receives from words is seen as being fundamentally of the same kind as the information one gets from interacting with the world, then there are good prospects for being able to integrate them. There may be uncertainty about how the world is according to an utterance, but this uncertainty will change if further information is provided lexically, from observation, or from knowledge of the context.

In the first two chapters of the next part (6 and 7), a formal model for the probabilistic structure of the meaning of vague predicates will be given. I shall propose that communication under conditions of uncertainty is possible because a certain amount of uncertainty can be eliminated, depending on one's communicative goal. In short, how precisely one needs to know how the world is according to an utterance depends upon what the purpose of that utterance is. In making this move, one of the benefits of taking a probabilistic approach to linguistic representation will be made apparent, namely, that this kind of context-sensitivity can be easily and naturally integrated into the semantics. That said, I will not be defending the view that certainty needs to be reached. This will mean that nothing in the combination of meaning and circumstance will, in itself, answer questions of truth for us. Hence, in chapter 8, we will return to truth, logic, and eventually, in chapter 9, the Sorites paradox.

PART III

SEMANTICS AND PRAGMATICS

6

Towards a Semantics

In chapter 1, we saw that many accounts of vagueness assume truth and truth-conditions to be core semantic notions for vague terms. In chapter 2, when communication and learning were pushed closer to the fore, we saw reason to doubt the prospects of a truth based strategy. In chapters 3, 4, and 5, a more positive view started to emerge, namely, that words carry information, something that can be simply decoded, but that this information is not like a proposition or a (Fregean) thought. Both from the perspective of communication (the kind of information we need to be able to decode) and learning (the kind of information that could survive the bottleneck), two related concepts started to replace the central place previously reserved for truth values and truth conditions. These were probability and uncertainty. The former arose when questions of formalisation were at issue. How should (semantic) information be formalised? The latter arose from far more human concerns. When words are used, we are left with uncertainty over (exactly) how the world is. This uncertainty can be put in terms of belief. What are we entitled to believe and with what force?

In this chapter and the next, both informal considerations pertaining to belief and uncertainty, and formal considerations pertaining to probability will be utilised to provide both a formal and an informal account of the semantics for vague expressions. This chapter will look at semantics, which, for our purposes, will mean what vague expressions encode. This will still

Author's note: Since this thesis was examined, I have substantially amended the views presented in this chapter. A more up-to-date presentation of my ideas can be found in *Towards a Probabilistic Semantics for Vague Adjectives*.

leave a great deal out, since what is encoded will be quite minimal. In the next chapter, on pragmatics, we will finally have recourse to consider truth and reference.

6.1 Uncertainty

There are two ways that uses of a vague term can give rise to uncertainty. The first concerns the world and the second concerns the word. Say that we are told that John is tall, or that Mary's car is green. Given that we are competent in English, and, given that we have no reason to expect what we have been told to be incorrect, it would be reasonable to believe certain things about the world on the basis of what we have been told. One feature of vague expressions is that our judgements about them seem to turn on differences in determinable features of the world.[1] We saw this reflected in many of the accounts discussed in chapter 1. When the semantics of a vague term is analysed, some less vague property or concept or feature is often cited in the analysis. For example, with respect to 'green', our judgements vary over the shades that things are; for 'tall', over the heights that things are; for 'bald', over how much (head) hair individuals have; for 'heap', over how many grains or stones, or whatever, there are. On being told that John is tall, or that Mary's car is green, one thing we might have beliefs about is, how tall John is (or his height), or what shade Mary's car is.[2]

In the following, I will represent the content of these beliefs with reference to measures that may not have any psychological reality. For example, I will talk about someone's beliefs about heights measured in, say, centimeters or inches. However, it need not be assumed that our doxastic representations of heights refer to centimeters or inches. For example, if we talk of Mary's beliefs about John's height as distributing over heights in centimeters, this is merely a convenient notation. It does not require that Mary can say, in centimeters, how tall John is (or might be).[3]

[1]By 'determinable', I simply mean in ways that can be described. I do not mean discriminable.

[2]Of course one way to describe what we might reasonably believe would be that John is tall or that Mary's car is green. Still, one might pursue what these beliefs say about John's height, or the shade of Mary's car.

[3]It is possible that by using precise values, I am introducing an artificial level of precision into the model. It could be that the way we cognitively represent the world is also vague. In which case, perhaps there should be a mapping from vague words to a mapping from a vague representational level to sharp properties in the world. I take it that by describing mappings between vague words and non-vague properties in the world, I am, at worst,

It would be unreasonable to believe that John is a specific height just on the basis of being told that he is tall, and it would be unreasonable to believe Mary's car to be a specific shade just on the basis of being told it is green. Herein lies the first kind of uncertainty:

(U1) Descriptions using vague expressions leave us uncertain about specific features/properties that objects in the world have.

U1 uncertainty is often, to some extent, eliminable in practice. We can, for example, go and look at John (maybe even measure him), or we can go and have a look at Mary's car. And yet, even if we are in command of these facts, uncertainty about how to use our language can still remain. For if John is, say, 180cm tall, or if Mary's car is an odd sort of turquoisey shade, we might feel thoroughly uncertain whether or not 'tall' would be an effective word to use to describe John or 'green' an effective word to use to describe Mary's car.

By using the term 'effective' here, I mean to appeal to a notion of success. What is effective (for some purpose) is what succeeds in doing something/bringing about the desired result. What is effective for describing John, or Mary's car is going to be an occasion sensitive and interest relative matter. What is effective for doing something in one situation may not be effective in another.

As we have seen, a lot of the literature on solutions to the sorites paradox take the only important criterion for what is effective to be what description would be TRUE. However, for many purposes, it might not be truth that we need to establish, or are interested in, for communication. For example, if our aim is to communicate which individual John is in a crowd, irrespective of whatever story about the truth-conditions for 'tall' are, one might be able to use 'tall' to identify John because he is significantly taller than those around him. In chapters 7 and 8, we will worry about true descriptions. In this chapter, I will continue to use the broader notion of *effective descriptions*. Hence we have a second kind of uncertainty:

(U2) Specific features/properties that objects in the world have can leave us uncertain about how to effectively describe them.

There are a few things to say about U1 and U2:

1. In principle, the two kinds of uncertainty could be independent. Suppose that there were exact rules governing language use such that

oversimplifying the matter by skipping over a mental/cognitive level of representation.

'tall' and 'green' had sharp boundaries. We might nonetheless be U1 uncertain, given a description of something as 'tall' or 'green', even though we would not be U2 uncertain about how to apply the terms (if we had enough knowledge about the world). Alternatively, imagine a term that only refers to some exact property or object. If we heard the term being used, we would have no U1 uncertainty about the object/property referred to. But if we lacked sufficient information about the world, we could still be U2 uncertain about whether to apply that term or not.

2. Yet, when we steer clear of these kinds of fictional examples, the two kinds of uncertainty are clearly related. If, in some situation, we were no more certain of applying 'tall' to 185cm John than to 190cm Bill, then, if we didn't know their heights, all else being equal, we should be no more certain that John (Bill) is 185cm in height when described as tall than 190cm. Furthermore, if we were no more certain that John was 185cm in height than 190cm, given that he has been described as tall, then, were we to discover his height, we should be no more certain of applying 'tall' to John at 185cm than we would be if he were 190cm.

3. Given 2, (where U1 similarities give rise to U2 similarities and *vice versa*), it would be reasonable to expect U1 differences to give rise to U2 differences and *vice versa*.

4. If we allow variations in U1 uncertainty, it can give rise to U2 uncertainty. Being told that John is tall may leave us uncertain about his height. We might be fairly certain that he is not 170cm, equally and moderately uncertain that he is 185cm or 190cm, but fairly highly uncertain about whether he is 180cm. Given 2 and 3, this translates into U2 uncertainty over the application of 'tall'. Say we are fairly certain that 'tall' effectively describes someone who is either 185 or 190cm in height, but fairly certain that it wouldn't effectively describe someone who was 170cm in height. We should be neither certain nor uncertain about some heights in between such as 180cm. We should also expect a range of U2 judgements to give rise to U1 uncertainty.

Two challenges arise for communicating with vague expressions. As a hearer, we must face uncertainty over what objects in the world are like, given the way they have been described. As speakers, we may, from time to time, face uncertainty over how to effectively describe things as being.

We started out by considering what it is reasonable to believe, given a certain description of the world. When describing the world oneself, the flip-side of this is what description would be reasonable to use (so as to be effective).

On the basis of what we have been told, what makes the beliefs that we have reasonable? The simple answer is: The information that those words carry. As per chapters 3 and 4, this information will be represented using probability theory. In particular, we will use conditional probability distributions.

6.2 Constituents

A semantics based on uncertainty (on the statistical, correlative notion of semantic information) will have to be able to attribute, to specific expressions, or perhaps even types of expression, a role in larger constructions in which they occur. Take our above examples: 'John is tall' and 'Mary's car is green'. Giving a semantics for the modifiers 'tall' and 'green' might then be taken to be a matter of accounting for what information they carry about some object. For example, 'x is green' might be modelled as making it more reasonable to believe that x, whatever x is, is some range of shades rather than others.

However, here a disparity with 'tall' arises. The x in 'x is tall' is vital for getting any idea about what it is reasonable to believe. If we are only told that something, anything, is tall, be it a mountain, a molehill, a mouse, or a millipede, there is no height it would be more reasonable to believe this thing is than others.[4] This changes as soon as 'tall' is predicated of an NP or is applied to a CN in an NP: The 'tall' in 'x is a tall man' seems to make it more reasonable to believe that x is some heights rather than others (1.9m rather than 1.6m, say). But this suggests that part of the information which contributes to our expectations of heights is coming from the CN/predicate 'man'.[5] This can be seen by substitution of CNs/predicates.[6] Compare how expectations of heights differ for 'x is a tall man', 'x is a tall molehill' and 'x

[4]This may not be quite right. There may be some priors we have or reasoning we could engage with that would make some heights more reasonable than others.

[5]I wish, again, to emphasise that the numbers here are purely convenient devices to describe a representation of heights. Actual beliefs do not need to be about centimeters.

[6]I do not mean to take, as assumed, a clear, semantically important division between CNs and adjectives. Below, I will treat words like 'tall' and 'green' as predicate modifiers and words like 'car' and 'man' as as predicates. In many cases, the traditional classifications of words as CNs and adjectives will overlap with these semantic divisions. Nothing I say rests on whether they all do.

is a tall mountain'.[7] The situation for green seems to be similar. CNs make a difference to expectations.[8] However, as noted 'x is green' does seem to carry some expectations of its own.

I take this to be a reason to see some differences in the semantics of various vague adjectives/modifiers. The modifier 'green', which seems to carry information on its own, will have a slightly different semantic shape to those modifiers (such as 'tall') that only seem to modify expectations based on information carried by nominal predicates. The details of this difference will be elaborated in §6.4.

What we have here is an argument for treating some modifiers as, in some sense, not really carrying information *per se*. Instead we can see them as encoding a modification on the information carried by a nominal predicate, or a modification on information that is contextually provided. Someone's being described as a woman gives us a rough expectation as to her height. This may be a very broad range, but we have some expectations nonetheless. Something's being described as a skyscraper gives us very different expectations as to its height. One might worry that this puts a lot of weight on information carried by nominal predicates. However, all that is being assumed is that learning to categorise and classify objects with such predicates, in part, amounts to developing expectations as to basic visual cues such as size, shape and shade.[9]

The common modifiers 'old', 'big' and 'long', just like 'tall', only seem to give us reasonable expectations when applied to a nominal predicate (or when themselves predicated of an NP in cases such as 'John is tall'). In turn, however, that suggests that nominal predicates such as 'man' carry information about what features men can reasonably be expected to have. Again, some tests (albeit ones based on linguistic intuitions) can be applied to get a grasp on what information this is. For example, at least for non-metaphorical uses, 'x is a long man' does not seem to make much sense/provide us with any reasonable expectations about x's length.[10] This suggests that 'man' does not carry information pertaining to length. This is not to say that modifiers such as 'green' are not restricted either. For example, synaesthesia aside, 'x has green ideas' is hard to understand if

[7]If readers are uncomfortable with these open sentences, substitute 'That' for 'x'.

[8]This point is similar to the commonly held idea that vague adjectives are context-sensitive or interpretable only relative to a comparison class (Cresswell 1974).

[9]See (Tenenbaum et al. 2011) for a good review of similar ideas in the literature.

[10]Length is not simply a euphemism for height. Trains, passageways, halls, to name a few, can have a great length without having a great height.

'green' relates to shade.[11]

I take this as a reason to treat nominal predicates as carrying information based on correlations between things that those common nouns are used to classify and the features that the things they are used to classify have. For example, it is because there is a stronger correlation between men and some heights than there is with others, that 'man' carries information about heights.[12] Adjectives will be modelled as modulating some specific aspect of the information that nominal predicates carry. For example, 'tall' will modify information relating to height carried by predicates like 'man'. Words like 'green' will do this too, but in a slightly different way to be elaborated below.

A possible worry needs to be flagged here.[13] In the literature on adjectives, it is common to appeal to a comparison class to capture the semantics (Cresswell 1974), (Kamp 1975), (Kennedy 1997), (Kennedy 2007), (Graff Fara 2000). However, it has been pointed out that common nouns do not always determine comparison classes.[14] For example, BMWs are not the comparison class below, despite featuring as part of a modified NP:

1. Kyle's car is an expensive BMW, though it's not expensive for a BMW. In fact it's the least expensive model they make. (Kennedy 2007)

The worry concerns whether the above analysis wrongly implies that the modified noun always indicates the comparison class. The analysis can, however be modified slightly. Rather than demand of some adjectives (tall, big, long etc.) that they simply modify a nominal predicate, we can instead require that there is some class of things to which it is being applied in any context (information about which can be modified by the adjective). This still leaves a difference with terms like 'red' which seem to carry some information (about, say, shades) independently of a lexically provided nominal predicate or a comparison class.[15] I will return to these cases in chapter 7.

[11]An interesting possibility is that the application of domain specific modifiers to NPs that do not carry information about that domain will generate metaphorical interpretations. Metaphor, humour and other such subjects are outside of the scope of this project.

[12]Recall, these correlations do not persist through space and time by accident.

[13]This was pointed out to me by an anonymous commentator for the ESSLLI 2013 workshop on Bayesian semantics.

[14]See (Kennedy 2007, §2.2) for a thorough overview.

[15]That is not to say that 'red' cannot relate to non-stereotypical shades (such as in 'red onion'). The analysis of a comparable case will be made in §6.5 .

6.3 Semantics: Preliminaries

6.3.1 Correlations and Situations

In the following, I will suggest one way to formalise U1 uncertainty (uncertainty over how the world is, given a description of it). I will borrow heavily from situation semantics (Barwise and Perry 1983), (Cooper 1996), (Devlin 2006).

In situation theory, meaning is relational (Devlin 2006, p. 602). The meaning of an expression is taken to be a relation between a *discourse situation*, a *speaker's connection* and a *described situation*. Bracketing speaker connections, meaning can, similarly, be treated as (cor)relational. A correlational account of meaning assigns meanings to expressions as correlations between discourse situations of a certain type, and described situations of a certain type.

In Situation Theory, situations support *infons*. Infons are traditionally conceived as what contribute to forming informational items (Devlin 2006), or as types of situations (Cooper 1996).[16] Another way to view them is as properties of situations (as opposed to properties of individuals). If a situation, s has some property (some infon), σ, then it is said that the situation supports the infon. That a situation supports an infon, in notation, is a situation theoretic proposition:

$$s \models \sigma$$

Here, situations are meant to be ways to classify objects and events in the world. They are not possible worlds. If a situation, in fact, has the property σ (is, in fact, of type σ/supports σ), then the proposition is true. However, we will not be seeking to associate linguistic expressions with propositions, nor will we be directly interested in particular situations and whether they support some infon.

Situations can be abstracted over to give *situation types* via situation type abstraction (Devlin 2006, p. 607):

[16] This way of viewing infons (as types) propagates through into situation theoretic approaches with richer type systems. See, for example (Cooper 2012). As well as from Cooper, the origins of this idea trace to Jon Barwise.

$$[\dot{s} \mid (\dot{s} \models \sigma)]$$

This means the type of situation in which σ obtains. The $\dot{s}$ denotes a *situation parameter*. Situation parameters, and other parameters, in situation theory, are placeholders for specific objects (Devlin 2006, p. 605). For our purposes, we can see situation parameters as variables. However, in the formalism, we will have need to use both situation parameters and variables. Variables will be provided values from a variable assignment function. Situation parameters will be assigned situations via an *anchoring* function.[17]

Abstracting over situations, as above, forms a *situation type*. However, strictly speaking, correlations do not hold between types of situations but between situations of some type:[18] If some concrete situation of some type arises, there will be some probability that a situation of another type will arise (or, arose, or will have arisen etc.). Correlations do not hold between individual (historical) situations, but between situations of a type.[19] This raises a tricky question as to how to represent such situations between which correlations hold. We do not want to use a specific situation to form a proposition ($s \models \sigma$), but nor will a situation type do (where situation types are functions from situations to propositions), since these are of the wrong type of entity to have a probability. As an approximation, I will make use of unanchored parameters. Unanchored parameters, as placeholders for specific situations, can be assigned a type (said to support an infon). The idea will be that if some specific situation is of type σ (one that supports σ), there will be a probability of another specific situation being of type/supporting an infon τ.

Hence probability values will be given to situation theoretic propositions formed with a situation parameter:

$$\dot{s} \models \sigma$$

[17]This will allow interpreted situation theoretic propositions formed from situation parameters to have probability values.

[18]This isn't necessarily the case if, instead of taking situation types as sets of situations, they are taken as objects themselves. This rich type-theoretic approach is taken in Type Theory with Records (TTR) (Cooper 2012). See below and chapter 7 for further discussion of the differences between this and my approach.

[19]This is not quite accurate. Probabilities of historical situations will be 1 or 0 depending on whether that situation is part of the world.

Unlike situation theoretic propositions, however, a parameterised proposition will not be true or false independently of the parameter being assigned a situation (by the anchoring function).

The picture of information built up in earlier chapters was that there are correlations between uses of words and certain ways for the world to be. U1 uncertainty (where descriptions using vague expressions leave us uncertain about specific features/properties that objects in the world have), will be captured using probabilities.

We will have situations of two kinds:

(i) *Discourse situations* are situations in which some specific type of discourse takes place. For example, we might have the discourse situation in which 'red' is used to classify some object, or, in which 'tall' is used to describe some person. Discourse situations will model the situations in which certain words are used (certain things are said).

(ii) *Described situations* are situations in which the world is some way. For example, we might have the described situation in which some object is some shade of colour, or, in which some person, say Mary, is 180cm in height. Described situations will model the kinds of ways the world can be.

An example of a proposition formed with a discourse situation parameter is one in which John is described as tall:[20]

$$\dot{d} \models \text{“tall”}(\mathtt{j})^{+}$$

There will be a probability of some described situation being of some type, given this one. For example, that John is some height:

$$\dot{s} \models \mathtt{height}{=}\mathbf{h}(\mathtt{j})^{+}$$

For different values of $\mathbf{h}$, these probabilities will form a distribution. Abstracting away from the contribution made by the name 'John', this distribution will reflect the extent to which uses of the modifier 'tall' to describe,

[20]Details of infon notation will be addressed below. In brief, “tall”$(\mathtt{j})^{+}$ means that the individual $\mathtt{j}$ was described using 'tall'.

say, a human male, correlates with human males being certain heights. The heights that an agent will entertain will be restricted by their learning process (namely, an approximation over the kinds of humans, males etc. that they have experienced). It will be supposed that a rational agent can only have distributions that sum to 1.

When someone makes a statement, the information carried will be captured as the probability distribution of a proposition formed with a described situation parameter, given a proposition formed with a discourse situation parameter. For example, the probability that John is 5ft in a described situation, given that he was described as tall in a discourse situation, will be (comparatively) very low. Other conditional probabilities will be greater for greater height properties John might have. A range of such probabilities (each with a different height property assigned to John in the described situation) will form a distribution (must sum to one).

Conditional probability values (of discourse situation propositions, given described situation propositions) will represent something like credence. Here, however, rather than just a degree of belief, we will be concerned with the degree to which it is reasonable to believe that some state of affairs obtains. The distinction being drawn here is between what, given some actual description, a hearer believes (credence), and what, in virtue of the information carried by the utterance of the words, one is entitled to believe. Recall that information, in the sense developed earlier, reflects the correlations between uses of words and states of affairs. This incorporates all uses, both representations and misrepresentations of the world. The degree to which it is reasonable that some state of affairs obtains reflects this.

So, if being told that John is tall makes it reasonable to believe to a degree of n that he is 180cm in height, this will be represented as assigning a probability value of n to the conditional probability of there being a situation of the type in which John is 180cm in height, given a situation of the type in which he is described as tall. However, it must be stressed that the numbers in the formalism are MODELLING extents to which we can reasonably believe states of affairs to obtain given the information we have been provided with. The value for n, against another larger number, represents a lower reasonable credence in a state of affairs than the higher number. Also, as mentioned earlier, centimetres and meters etc. are merely convenient ways to talk about heights.[21]

[21] However, of strict importance is that values for reasonable beliefs form a distribution (sum to 1).

Types and Situation Types

A more recent approach to situation theoretic semantics is systematically laid out in (Cooper 2012) which adopts a rich proof theoretic type theory. Rather than treating types as sets of objects in a domain, this approach treats types as objects (in a domain of types). Situations/events then act as proofs of propositions. In what is to follow, I will adopt a fairly simple type theory with an, essentially, model theoretic semantics.[22] This simplification will make it easier to focus on the probabilistic element in my account. A proof theoretic approach with a richer type theory could, in principle, adopt the core ideas in the proposal that I will make. Indeed, a probabilistic approach comparable to my own has just been proposed in (Cooper et al. 2013).[23]

Properties

It is important to note that we are not assuming that there is anything like the property *green* or the property *tall* that individuals have or lack. We will be assuming that individuals can have the property of being some height (this will just be the height that they are). We will also be assuming that certain objects have some hue or shade. These heights and shade properties will be mentioned in the formalism as properties of infons.[24]

This does not require that properties are not vague. I wish to stay neutral on the issue of whether or not being some shade or being 180cm in height are vague properties (1.2.1). However, for modelling simplicity, I assume that properties are sharp.

6.3.2 Definitions and Axioms

Types

The type system that will be used will have three basic types. Rather than the traditional type t (for truth value), we will have type p (for probability). Type e will be the standard type for individuals (entities). Type s will be the type for situation. Basic types can be used to form functional types in the normal way:

[22] The hedging here is included because, although I assume a domain of situations (as a set), situations are best not viewed as a set.

[23] A difference between this, my own, and similar proposals will be made at the end of chapter 7.

[24] Infons are properties of situations. Properties of objects are properties of infons.

Definition: Types

Basic types	$\{e, p, s\}$
Functional types	If a, b are types, then $\langle a, b\rangle$ is a type.

Expressions

This formalism brings together notions from both probability theory and situation theory. As such, we will need to define how various parts of them interact. The expressions of the formalism will be as follows:

Definition: Vocabulary

Variables (VAR):	infinite set of variables for all types.
Constants (CON):	a possibly empty set of constants for every type
Situation parameters (PAR):	$\dot{s}_1, ..., \dot{d}_1, ...$
Supports:	$\models$
Conditional on:	$\vert$
Infons (with polarities):	σ^+, σ^-, τ^+, τ^-
Infon Connectives:	$\wedge_{in}, \vee_{in}$
Probability Connectives:	$\wedge_{pr}, \neg_{pr}, \vee_{pr}$
Mathematical Operators:	$\times, +, -$
Numerical Constants:	$\mathbf{C}, \mathbf{C}'$ of type $\langle p\rangle$
Brackets:	(,)

Where necessary, types of expressions will be given as subscripts. For example, a variable, S, of type $\langle e, s\rangle$ will be written $S_{\langle e,s\rangle}$. Of particular note are the two versions of each connective. These will be discussed below.

A further few remarks on infons are needed at this juncture. I assume a class of properties that, along with individual expressions, and a polarity, form infons. Positive polarities on infons indicate that any situation that supports the infon is of that type. A negative polarity would indicate that the situation is of the negative type.[25] Barwise puts it as follows (he uses the terms 'state of affairs' and 'state' instead of 'infon'):

> "Note that for a basic state of affairs σ and its dual σ' (the state with the opposite polarity but everything else the same), either $\models \sigma$ or $\models \sigma'$ (and not both). However, for a particular situation, it

[25]The '+' and '−' are equivalent to '1' and '0' in (Devlin 2006) and 'yes' and 'no' in (Barwise and Perry 1983). In what is to follow, readers should assume a positive polarity unless stated otherwise.

may well not be the case that either $s \models \sigma$ or $s \models \sigma'$, since s may not determine which of these is the fact of the matter." (Barwise 1989, p. 185)

Infons can be understood as types of situations (Cooper 1996), or perhaps as properties of situations. As such, they are simply ways for things to be (or not to be). The class of infons will be formed of all possible combinations of properties and objects in our ontology (plus a polarity). Infons are used to classify situations (to say of situations that they are some way). An infon is a property provided with an entity as an argument. Properties and infons are not themselves fully expressible within this system.[26]

The well-formed expressions of the language are defined (where $\mathrm{WE_a}$ denotes well formed expressions of type a):

Definition: Well-Formed Expressions:[27]

(i) If $\alpha \in \mathrm{VAR_a}$, then $\alpha \in \mathrm{WE_a}$,

(ii) If $\alpha \in \mathrm{CON_a}$, then $\alpha \in \mathrm{WE_a}$,

(iii) If $\alpha \in \mathrm{PAR_s}$, then $\alpha \in \mathrm{WE_s}$,

(iv) If $\alpha \in \mathrm{WE_{\langle a,b\rangle}}$ and $\beta \in \mathrm{WE_a}$, then $\alpha(\beta) \in \mathrm{WE_b}$,

(v) If $\alpha \in \mathrm{WE_b}$, and $\beta \in \mathrm{WE_a}$, then $\lambda\beta.\alpha \in \mathrm{WE_{\langle a,b\rangle}}$,

(vi) If $\dot{s} \in \mathrm{WE_s}$, and if σ is an infon, then $\dot{s} \models \sigma \in \mathrm{WE_p}$,

(vii) If $\phi, \psi \in \mathrm{WE_p}$, then $\phi \mid \psi \in \mathrm{WE_p}$,

(viii) If $\phi, \psi \in \mathrm{WE_p}$, then:

  (a) $\neg_{pr}\phi \in \mathrm{WE_p}$,

  (b) $\phi \wedge_{pr} \psi \in \mathrm{WE_p}$,

  (c) $\phi \vee_{pr} \psi \in \mathrm{WE_p}$,

(ix) If $\dot{s} \models \sigma, \dot{s} \models \tau \in \mathrm{WE_p}$, then:

  (a) $\dot{s} \models \sigma \wedge_{in} \tau \in \mathrm{WE_p}$,

[26] This slight fudge is the upshot of the tension created by my sticking to a simple (denotational) system of types whilst still using some of the richer vocabulary of situation theory. That properties and infons do not fit into a system of simple types could suggest that a richer notion of types should be incorporated into the system (see comments on (Cooper 2012) above).

[27] Where a, b are types.

(b) $\dot{s} \models \sigma \vee_{in} \tau \in \mathrm{WE_p}$,

(x) If $\phi, \psi \in WE_p$, then

(a) $\phi \times \psi \in \mathrm{WE_p}$,

(b) $\phi + \psi \in \mathrm{WE_p}$,

(c) $\phi - \psi \in \mathrm{WE_p}$,

Clauses (i), (ii), (iv) and (v) are standard (see e.g. (Carpenter 1997)). Clause (iii) states that situation parameters are well formed. Clause (vi) introduces the syntax for situation theoretic propositions (formed with situation parameters). Clause (vii) gives what will eventually be interpreted as a conditional probability. The connectives in clauses (viii) and (ix) will be discussed in greater detail below. Clause (x) allows for expressions of type p (probability) to be linked by mathematical connectives.

As mentioned briefly earlier, there are two kinds of situation parameters that we will use: *described situations* ($\dot{s}$) and *discourse situations* ($\dot{d}$). Both of these can support infons. However, for descriptive situation types, we will be interested in how the world is with respect to things like the shades things are (for colour terms), the heights things are (for terms like 'tall') and the numbers of hairs things have (for terms like 'bald'). In discourse situation types, what will be described is the production of some utterance.

The infons we will be concerned with are likewise of two kinds. Infons for descriptive situations will describe/be about, say, what height someone is. Immediately below I specify the infon, and below that, I give a parameterised proposition stating that some situation supports that infon:

$\texttt{height=180cm(j)}^{+}$

$\dot{s} \models \texttt{height=180cm(j)}^{+}$

The latter is a situation theoretic proposition. It states that, in $\dot{s}$, the individual denoted by `j` (say, John) is 180 cm in height. Infons for descriptive situation types will typically relate to what has been said. For example, below, I specify a discourse infon, followed by the statement that a discourse situation supports that infon:

$$\text{“}\texttt{tall}\text{”}(\texttt{j})^{+}$$

$$\dot{d} \models \text{“}\texttt{tall}\text{”}(\texttt{j})^{+}$$

The latter is a proposition containing a situation parameter. The proposition states that, in $\dot{d}$, the individual denoted by $\texttt{j}$ (say, John) has been described as tall. This is a simplification, however. In actual fact, there ought to be some abstraction over speakers and times etc. I suppress these details here.

For either of these situations (descriptive and discourse), many actual situations could be of those types (types where John is 180 cm in height, or where he is described using 'tall'). Parameters will get values/be interpreted via an anchoring function.

Interpretations and Domains

Rather than a standard interpretation function ($[\![\cdot]\!]$), the interpretation function we shall use will be a probabilistic one: $p[\cdot]$.

I assume a domain for each basic type. If a is a basic type:

$$p[\alpha_{\langle a \rangle}] \in Dom_{\langle a \rangle}$$

Functionally typed expressions will be interpreted as functions from basic domains:

$$p[\alpha_{\langle a,b \rangle}] \in Dom_{\langle b \rangle}^{Dom_{\langle a \rangle}}$$

The type e domain will be a set of individuals. The domain of type p will be the range $[0, 1]$. Situations are being taken as primitive. We can, for the sake of the formalism, assume a domain of situations. However, in the spirit of situation theory, situations should be taken to be our ways of conceptualising and carving up the world into parts.

As well as the interpretation function, we will have a variable assignment function g, which assigns an appropriate member of the domain to every variable.

In principle, we should also have an anchoring function h, however, we will see that this will not be used. Of particular importance is that the anchoring function will not be used in the interpretations of expressions containing situation parameters. This is intended to capture the idea that correlations hold between situations which have certain properties in common, but not between specific situations themselves.[28]

Definition: Denotations

(1) $p[x]^g = g(x)$ if $x \in VAR$,

(2) $p[\mathsf{c}]^g = p[\mathsf{c}]$ if $\mathsf{c} \in CON$,

(3) $p[\alpha(\beta)]^g = p[\alpha]^g(p[\beta]^g)$,

(4) $p[\lambda x.\alpha]^g = f$ such that where x is of type a and $c \in Dom_a$, $f(c) = p[\alpha]^{g[x:=c]}$,

(5) Maths Operators:

(i) $p[\phi \times \psi]^g = p[\phi]^g \times p[\psi]^g$ (multiplication),

(ii) $p[\phi + \psi]^g = p[\phi]^g + p[\psi]^g$ (addition),

(iii) $p[\phi - \psi]^g = p[\phi]^g - p[\psi]^g$ (subtraction),

(6) Probabilistic Connectives:

(i) $p[\neg_{pr}\phi]^g = 1 - p[\phi]^g$,

(ii) $p[\phi \wedge_{pr} \psi]^g = p[\phi]^g \times p[\psi \mid \phi]^g$,

(iii) $p[\phi \vee_{pr} \psi]^g = p[\phi]^g + p[\psi]^g - p[\phi \wedge \psi]^g$,

The first four clauses are fairly standard.[29] Clause (5) simply states that maths operators in the object language are treated as standard when the interpretations of their operands are in the range $[0, 1]$. Clause (6) simply states the standard definitions of probabilistic connectives (Kolmogorov 1950).

Axioms

There are both probabilistic connectives and infon connectives. The former connect expressions of type p, the latter connect infons. Where we are

[28]However, one can estimate a correlation by observing specific situations.
[29]c.f. (Carpenter 1997).

referring to the same descriptive situation or the same discourse situation, there is a connection between these two. Given the statement that a situation is of some type and the same situation is of another type ($\dot{s} \models \sigma \wedge_{pr} \dot{s} \models \tau$), we can conclude that the same situation is of both types ($\dot{s} \models \sigma \wedge_{in} \tau$). Given the statement that a situation is of a conjunctive type, we can conclude that the situation is of the type of both types in the conjunction. Likewise for disjunction:

$$\dot{s} \models \sigma \wedge_{in} \tau \Leftrightarrow (\dot{s} \models \sigma) \wedge_{pr} (\dot{s} \models \tau)$$
$$\dot{s} \models \sigma \vee_{in} \tau \Leftrightarrow (\dot{s} \models \sigma) \vee_{pr} (\dot{s} \models \tau)$$

Spaces and Priors

In order for these axioms and the probability function to be well defined, we must define a probability space. Probabilities will distribute over situation theoretic propositions (statements that propositions support/don't support infons). Were we to include an anchoring function in interpretations, then all propositions would receive a value of 0 or 1. Without an anchoring function, the value will be in the range $[0, 1]$ where the value indicates the probability of some situation being of the type specified by the infon. For all non-conditional values, I assume that probabilities distribute evenly. For example:

$$p[\dot{s} \models \mathtt{height}=\mathtt{h}(x)]^g = \frac{1}{\text{number of values of } \mathtt{h}}$$

In practice, the range of situations and infons will be highly constrained. For described situations, constraints will come out of the semantic learning process, as well as the goals and purposes of the speakers. For example, certain ranges of shade/hue properties will constrain the range of described situations for colour terms in general, but further constraints may come from what objects are salient to speaker and hearer in the situation of utterance (more details of this will be given in chapter 7).

For discourse situations, I simply assume a space of two infons, a positively and a negatively polarised infon formed with the same property:

$$p[\dot{d} \models \tau^+]^g \;=\; 0.5 \;=\; p[\dot{d} \models \tau^-]^g$$

These discourse situations relate to ones in which a predication is said to apply to an object, and ones in which this is stated not to be the case.[30] However, when multiple terms are used, these space of possibilities may be bigger and so values may be lower (but nonetheless equal).

These simplifying assumptions are being made because unconditional probabilities are not being assumed to play an important role in natural language semantics. That is why all described situation propositions (contextually restricted) will receive flat values and all discourse situation propositions for an utterance will also receive flat values. All of the work will be be done with conditional values. That is because it will be conditional values that will indicate the strengths of correlations between types of described and discourse situations of the relevant types.

Linguistic Meaning

The linguistic meaning of a word will be modelled as a correlation between unanchored parameterised propositions. For example, this will model the correlation between John being some height and John being is described as 'tall'. However, these values will be determined by more general correlations in which, say, people are described as tall and where people are some height or other (the details of this will emerge in the compositional semantics). The standard way we will represent declarative sentences will therefore be as a conditional probability formed with a described situation proposition and a discourse situation proposition:

$$\dot{s} \models \sigma \mid \dot{d} \models \tau$$

The interpretation of this will never involve the anchoring function.[31] It will be evaluated only by the interpretation function and a variable assignment:

$$p[\dot{s} \models \sigma \mid \dot{d} \models \tau]^g$$

Which will be in the range $[0, 1]$.

[30] For propositions that contain neither τ^+ nor τ^-, the situation may or may not support the infon. However, for simplicity, this possibility is being assumed not to hold.

[31] If it did, values would be only 0 or 1.

Representationalism

I have not mentioned a debate within formal semantics and linguistics over representationalism. This debate should not be confused with the debate in the philosophies of mind and perception. The semantics debate got going with the publication of work on Discourse Representation Theory (DRT) (for a detailed introduction, see (Kamp and Reyle 1993)). DRT questioned Montague's claim that the representational language of a formalism is dispensable. On a Montague semantics, natural languages are mapped to expressions in an intensional lambda calculus. Those expressions are then interpreted in a model. In principle, however, the expressions in the natural language could be mapped directly into the model. In DRT, it was claimed that to account for the treatment of some expressions (such as discourse anaphora), the representational (interpretational) level is not dispensable. Non-representational, dynamic accounts were developed in response (see for example (Groenendijk and Stokhof 1991)).

The debate was never entirely settled, but I do not need to take up a position. The presentation of the semantics above is in the spirit of Montague, albeit without the grammar that maps strings to expressions in the lambda calculus. This should not be taken as a commitment to anti-representationalism.

6.3.3 A Sentential Example

Before we enter into the role of modifiers in the account, I will look at an example of the treatment of a whole declarative sentence. Declarative sentences will, with a little artificiality, be interpreted as the probability of some situation being of a particular type, given a discourse situation of the type in which the sentence is used.

An example: Jess and Daniel are discussing what colour curtains to put up. Fortunately, both have before them, the curtain manufacturer's book of fabric swatches. Unfortunately, they have been forced to make their decision over the phone. Fortunately, each swatch has an identifying number on the back. Unfortunately, these long unwieldy identifiers aren't very practical for telephone conversation. As a result, Jess and Daniel stick to regular colour terms and modifiers in their utterances. Suppose that there are 64 swatches. Of these, suppose that 8 are more or less describable as being green. If Daniel says to Jess, "I like the green one", he is not being terribly informative, but he is, nonetheless giving Jess some information. He is, at least, cutting down the number of swatches it could be from 64 to more or less 8. As an idealized case, we could define 'green*' as a classifier that is only ever applied to the shades those 8 swatches are. Further assume that 'green*' is not a graded term and so applies equally to all of those shades. In that case, we could represent the information carried by Daniel's utterance with what it would be reasonable for Jess to believe regarding the swatch Daniel has professed to like. Again simplifying, we can assume that only one swatch is the swatch Daniel likes. In this case, Jess should believe with equal force, that each of the 8 green* swatches is the one Daniel likes. She should not put any weight on any of the other 56. So, Jess would have weak beliefs (with a strength of $\frac{1}{8}$) about the 8 green* swatches, and beliefs with strength 0 about the others, with respect to which swatch Daniel likes. We can represent what it would be reasonable for Jess to believe (and how strongly) as the extent to which there is a correlation between the discourse situation and a described situation. Over multiple described situations, this can be represented as a probability distribution:

'Daniel likes the green* swatch'

n	$p[\dot{s} \models \mathtt{like}(\mathtt{Daniel})(\mathtt{swatch}_n)^+ \mid \dot{d} \models \text{“Daniel likes the green* swatch”}(\mathtt{swatch}_n)^+]$
1	0.125
2	0.125
3	0.125
4	0.125
5	0.125
6	0.125
7	0.125
8	0.125

Distributions sum to 1. In the above, the 56 swatches which get a value of 0 are not represented. It may seem odd to think of there being any correlation between a situation in which Daniel professes to like the green* swatch and situations where Daniel likes some particular swatch. Daniel may never have expressed a swatch preference before. However, correlation will be between more basic elements of language (uses of 'green*' with things of particular shades). How these fuller distributions are constructed compositionally will be addressed below.[32]

[32]In all of what is to follow, I associate names with simple individual constants. However, calling an individual 'Daniel' might also carry information. To incorporate this into the

Now we can start to incorporate gradience.[33] Suppose that four of the eight swatches are comfortably describable as 'green', but that four of them are more borderline cases. Borderline cases (blueish green shades and yellowish-green shades) will be less strongly correlated with descriptions as 'green' than more central cases. This means that Jess has more reason to believe Daniel likes any of the four more central swatches than any of the other four. Now we could again represent what it would be reasonable for Jess to believe as a probability distribution. Low values are given to the borderline swatches (numbers 1, 2, 7 and 8). Equal credence applies to the central cases (numbers 3, 4, 5 and 6).

n	$p[\dot{s} \models \texttt{like}(\texttt{Daniel})(\texttt{swatch}_n)^+ \mid \dot{d} \models \text{"Daniel likes the green swatch"}(\texttt{swatch}_n)^+]$
1	0.05
2	0.05
3	0.2
4	0.2
5	0.2
6	0.2
7	0.05
8	0.05

Before we look at how such results could be built up compositionally, there are a couple of points to make about the semantics constructed so far. First, there has been no mention of truth in the model. Indeed, the standard semantic type t with domain $\{0, 1\}$ has been replaced with type p (for probability), with domain $[0, 1]$. We can, of course, ask, given the different situation types (with different values for n), which would make the utterance true? Yet, notice that nothing in the above semantics determines an answer to this question. Say that Daniel likes swatch$_3$. His utterance does not make this evident to Jess (or, at least, does not make it exceptionally reasonable for her to believe). Daniel has not effectively described the actual situation in that Jess has no more reason to believe that any single swatch is the one Daniel likes above any other. There are four that Jess has more reason to believe he likes (given what he said). However, even if he had liked swatch$_1$ (which received a value of only 0.05), what is encoded by his words

account, one could add, for example, a 'called-N' name predicate as well as an individual constant.

[33]Here, incorporating gradience could be taken to be the beginning of an incorporation of vagueness.

still does not determine that what he said was true (or false).

We can now begin to see how vagueness might emerge from the account. It arises at the periphery of use and learning. I will return to this at the end of the chapter. In brief, some shades of colour will have the same probability of being the shade some object is, given it has been described as green, as being shade the object is, given that it has been described as not-green. This tension arises from the information carried by the term 'green'. In this sense, the meaning of 'green' incorporates borderline cases.

I now move to a more detailed analysis of the semantics of individual terms, including vague adjectives/predicate modifiers.

6.4 Semantics: Terms

Predicates

Since nominal predicates were argued to be important for the semantics of vague adjectives/modifiers (as the things bearing the information that they modify), we will now turn to them. As suggested in §6.2, nominal predicates may carry a lot of different information. Our basic semantic representation for such predicates will incorporate an argument the range of which will be a selection of a type of information (such as a range of heights). This information type will be supplied either by context, or by a modifier, such as an adjective. In a departure from a simple view of predicates (that take just an object as an argument), nominal predicates will be modelled as a function from properties to a function from an individual to a proposition, which is to say that they will be a function from properties to properties ($\langle e, p\rangle, \langle e, p\rangle$). The logical structure of nominal expressions will be such that two entities in the above will be the same. Put another way, singular nominal predicates are assumed to require updating in context with respect to the aspect of the information they carry. Then, when provided with an individual we will get a value for the probability of that individual having that property, given the information provided. The schema for 'man' (and with appropriate substitution, other predicates) is:

$$\text{Man}: \quad \lambda S.\lambda x.\ \ (S(x) \mid \dot{d} \models \text{“}\texttt{man}\text{”}(x)^{+})_{\langle e,p\rangle,\langle e,p\rangle}$$

x, y, z	:=	Variables ranging over individuals. $(Type : \langle e \rangle)$
S, R	:=	Variables ranging over abstractions of individuals from propositions i.e. $\lambda y.\ \dot{s} \models (\texttt{height} = \texttt{180cm}(y)^{+})_{\langle e,p \rangle}$ (the type of situation in which y is 180 cm in height)

To get the information carried by 'is a man' with respect to the having of some property, we need to provide a property (an abstracted situation, or, in Devlin's terminology, an *object type*). Such a property (or, as we shall see, range of properties) may be provided by the context. Below, however, we will see how part of the semantic contribution of predicate modifiers is to provide such a property (such a range of properties).

Because such nominal predicates are functions from properties of described situations to a function from individuals to a proposition, a worry over tractability and learnability arises. If an agent needs to learn all distributions that arise from interpreting 'x is a man' constructions, this would not be a tractable learning task. However, probabilistic learning affords a solution. We can assume that agents begin classifier learning tasks with flat prior distributions for all collections of properties.[34] On being exposed to uses of a classifier, distributions will be adjusted for just those properties the objects are perceived to have. In this sense, if one has no reasonable expectations regarding some range of situations, given what has been said, simply because no one situation type is more plausible than another, then it is possible that the classifier used does not carry information about (the properties in) those situations.

One property a man might have is of being a certain height. For example, he might be 180cm in height in a situation of that type:

$$\lambda y.(\dot{s} \models \texttt{height=180cm}(y)^{+})$$

This kind of property can be applied to our representation of 'is a man':

[34]This will apply, at least, to the very early stages of learning. I do not wish to rule out that background knowledge will be brought to bear later on and this could be represented by a difference in priors.

$$
\begin{aligned}
&\lambda S.\lambda x.(S(x) \mid \dot{d} \models \text{“}\mathtt{man}\text{”}(x)^+) \quad \lambda y.(\dot{s} \models \mathtt{height{=}180cm}(y)^+) \\
&\Rightarrow \lambda x.(\lambda y.(\dot{s} \models \mathtt{height{=}180cm}(y)^+)(x) \mid \dot{d} \models \text{“}\mathtt{man}\text{”}(x)^+) \\
&\Rightarrow \lambda x.(\dot{s} \models \mathtt{height{=}180cm}(x)^+ \mid \dot{d} \models \text{“}\mathtt{man}\text{”}(x)^+)
\end{aligned}
$$

This is now in the right shape to take an individual as an argument which will be described as a man in $\dot{d}$ and assigned the height property in $\dot{s}$. Assuming (as a simplification) that names refer and carry no other information, this means that 'John is a man' can be derived and represented, with respect to being 180cm in height, as follows:[35]

$$
\begin{aligned}
&\lambda x.(\dot{s} \models \mathtt{height{=}180cm}(x)^+ \mid \dot{d} \models \text{“}\mathtt{man}\text{”}(x)^+) \quad (\mathtt{j}) \\
&\Rightarrow (\dot{s} \models \mathtt{height{=}180cm}(\mathtt{j})^+ \mid \dot{d} \models \text{“}\mathtt{man}\text{”}(\mathtt{j})^+)
\end{aligned}
$$

The interpretation of this will be a probability value that reflects, with respect to John having some height, the probability of John having that height, given that he is has been described as a man.

Of course, 'John is a man' will carry more information than this about John, but let us focus on information carried about his height. We can consider the range of heights John might be (given that he has been described as a man). There will be a rationality constraint on the values 'John is a man' receives with respect to John's height. The values, must sum to 1. As a simplification, we may think in discrete values, although a more accurate representation would be a continuous function.

Put another way, providing different arguments of height properties for the formula above will generate a probability distribution. Below is how one might look. I simplify slightly by taking ranges of heights:

[35] I ignore the contribution of the indefinite article and treat this as a simple predication `man(john)`. I leave the modelling of quantification in this system for future research.

h	$p[\dot{s} \models \texttt{height=}\mathbf{h}\texttt{cm(j)}^{+} \mid \dot{d} \models \text{“}\texttt{man}\text{”}\texttt{(j)}^{+}]^{g}$
$h < 150$	0.01
$150 \leqslant h < 155$	0.04
$155 \leqslant h < 160$	0.08
$160 \leqslant h < 165$	0.12
$165 \leqslant h < 170$	0.15
$170 \leqslant h < 175$	0.20
$175 \leqslant h < 180$	0.15
$180 \leqslant h < 185$	0.12
$185 \leqslant h < 190$	0.08
$190 \leqslant h < 195$	0.04
$h > 195$	0.01

This reflects that describing John as a man, in part, conveys information about his height: Namely, that it is fairly probable that he is around average height and highly probable that he is neither far below nor far above average height (though neither is inconsistent with the information carried). This relates to the assumption that learning the classifier ‘man’, in part, involves learning, very approximately, in what kind of height range men tend to come in. This needn’t be in centimetres or feet and inches, or whatever. It need only be internalised in such a way as to aid, to some extent, the ability to identify and classify whether objects in one’s environment are men.

Tall

As a predicate modifier, ‘tall’, when applied to a nominal predicate increases our (reasonable) expectations of the entity (that predicate is applied to) having a greater height. ‘Tall’ will therefore be modelled as having two jobs to do when applied to a common noun like ‘man’. One: it will pick out the information carried by ‘man’ with respect to height. Two: it will be a function on the probability distribution that ‘man’ generates with respect to heights (it will make taller heights more probable and shorter heights less probable). In terms of a distribution curve, for a graph with probabilities on its y-axis and heights on its x-axis, it would shift the whole curve along the x-axis in the direction of greater heights. The representation of ‘tall’ is given below, along with a derivation for ‘tall man’.

tall:

$$\lambda P.\lambda y.\mathbf{C}\times\ f_{\text{tall}}(\ (P(y))\ (\lambda z.\ \dot{s}\models(\texttt{height=h})(z)^{+}))$$

man:

$$\lambda S.\lambda x.\ \ (S(x)\mid\dot{d}\models\text{“}\texttt{man}\text{”}(x)^{+})$$

tall man:

$$\begin{aligned}
&\lambda P.\lambda y.\mathbf{C}\times\ f_{\text{tall}}(\ (P(y))\ (\lambda z.\ \dot{s}\models(\texttt{height=h})(z)^{+}))\\
&\qquad\qquad\qquad\qquad\qquad\lambda S.\lambda x.\ (S(x)\mid\dot{d}\models\text{“}\texttt{man}\text{”}(x)^{+})\\
&\Rightarrow\lambda y.\mathbf{C}\times\ f_{\text{tall}}(\ (\lambda S.\lambda x.\ (S(x)\mid\dot{d}\models\text{“}\texttt{man}\text{”}(x))(y))\ (\lambda z.\ \dot{s}\models(\texttt{height=h})(z)^{+}))\\
&\Rightarrow\lambda y.\mathbf{C}\times\ f_{\text{tall}}(\ \lambda x.\ ((\lambda z.\ \dot{s}\models(\texttt{height=h})(z)^{+})(x)\mid\dot{d}\models\text{“}\texttt{man}\text{”}(x)^{+})(y))\\
&\Rightarrow\lambda y.\mathbf{C}\times\ f_{\text{tall}}(\ \lambda x.\ (\dot{s}\models(\texttt{height=h})(x)^{+}\mid\dot{d}\models\text{“}\texttt{man}\text{”}(x)^{+})(y))\\
&\quad\Rightarrow\ \lambda y.\mathbf{C}\times\ f_{\text{tall}}(\ \dot{s}\models(\texttt{height=h})(y)^{+}\mid\dot{d}\models\text{“}\texttt{man}\text{”}(y)^{+})
\end{aligned}$$

$P :=$	Variable ranging over representations for predicates (functions from situation properties to a function from individuals to $[0,1]$). $Type:\ \langle\langle e,p\rangle,\langle e,p\rangle\rangle$
$\mathbf{C} :=$	Normalising constant.
$f_{tall} :=$	Expression of type $\langle p,p\rangle$

The interpretation of the result is a function from individuals to a probability distribution for heights of that individual, given their description as a tall man. The '**C**' is a normalising constant. Whatever function on distributions f_{tall} is, **C** will normalise that distribution (make it sum to 1).[36] One possible (although unrealistic) interpretation for f_{tall} could be that it has the effect of shifting the distribution for 'man' up by 20cm or so (and stretching it vertically).[37] Using the above toy distribution for 'John is a man', we can get a toy distribution for 'John is a tall man'.[38] Where:

$$\dot{s}\models(\texttt{height=h})(\texttt{j})^{+}\mid\dot{d}\models\text{“}\texttt{man}\text{”}(\texttt{j})^{+}$$

Is abbreviated below as Φ:

[36] The value for **C** will be 1 over the sum of the modified distribution.
[37] This could also be described by adjusting values of parameters on a Gaussian function.
[38] I assume that 0.01 is the arbitrarily small value

h	$p[\Phi]$	$p[f_{\text{tall}}(\Phi)]$	$p[\mathbf{C} \times f_{\text{tall}}(\Phi)]$
$h < 150$	0.01	0.01	0.01
$150 \leqslant h < 155$	0.04	0.01	0.01
$155 \leqslant h < 160$	0.08	0.01	0.01
$160 \leqslant h < 165$	0.12	0.01	0.01
$165 \leqslant h < 170$	0.15	0.01	0.01
$170 \leqslant h < 175$	0.20	0.02	0.02
$175 \leqslant h < 180$	0.15	0.06	0.06
$180 \leqslant h < 185$	0.12	0.10	0.11
$185 \leqslant h < 190$	0.08	0.20	0.22
$190 \leqslant h < 195$	0.04	0.30	0.32
$h > 195$	0.01	0.20	0.22

Simply translated, being told that John is a man may carry all sorts of information. Some of this is that he is more likely than not to be within some margin of average height. Being told that he is tall carries the information that it is highly probable that John is over some height (represented here as 180cm). 'Tall' selects a particular kind of information carried by its common noun (namely information about height), and it amplifies expectations for the upper bounds.

Importantly, after normalisation, we are simply left with a probability distribution (in the final column). Any of those values will be able to enter into formulas for further manipulation if need be. It will be convenient to mark that normalised distributions are mere distributions in the syntax of the formalism. This can amount to a dropping of the f and $\mathbf{C}$, but keeping a record of what the normalised distribution is over. Hence, the interpretation of formulas like:

$$\mathbf{C}\times\ f_{\text{tall}}(\dot{s} \models (\texttt{height=h})(\texttt{j})^{+} \mid \dot{d} \models \text{``}\texttt{man}\text{''}(\texttt{j})^{+})$$

will be a probability distribution over heights, which can be rewritten with an updated discourse situation:

$$\dot{s} \models (\texttt{height=h})(\texttt{j})^{+} \mid \dot{d} \models \text{``}\texttt{tall(man)}\text{''}(\texttt{j})^{+}$$

However, in doing so we must be aware that we have increased the probability space of discourse situations. Now there are four possible discourse situation types since John might or might not be described as a man and might or might not be described as tall.

Green

As noted in §6.2, modifiers like 'green' will work differently from 'tall': simply knowing that something (anything) is green WILL give reasonable expectations about what it is like. The way this will be modelled is that 'green' will be also be a function on probability distributions, but it will work differently from 'tall'. Whereas 'tall' was informally characterised as taking a distribution over heights and shifting it up (so that greater heights receive higher values and lower heights, lower values), 'green' will contribute something more stable to a distribution. It will select shades or hues as properties to distribute over, but it will have the effect of weighting the distribution towards certain shades. For each shade that an object might be, 'green' provides a weighting over our expectations for what is being referred to as being one of those shades.

Formally, 'green' will look the same as 'tall'. The difference will be in the functions f_{tall} and f_{green} (see below). Immediately below is the derivation for 'green car'. Again **C** is a normalising constant.

green:

$$\lambda P.\lambda y.\mathbf{C}\times\ f_{\text{green}}((\ (P(y))\ (\lambda z.\ \dot{s} \models (\texttt{shade=c})(z)^{+})))$$

car:

$$\lambda S.\lambda x.\ (S(x) \mid \dot{d} \models \text{“}\texttt{car}\text{”}(x)^{+})$$

green car:

$$\lambda P.\lambda y.\mathbf{C}\times\ f_{\text{green}}(\ (P(y))\ (\lambda z.\ \dot{s} \models (\texttt{shade=c})(z)^{+}))$$
$$\lambda S.\lambda x.\ (S(x) \mid \dot{d} \models \text{“}\texttt{car}\text{”}(x)^{+})$$
$$\Rightarrow \lambda y.\mathbf{C}\times\ f_{\text{green}}(\ (\lambda S.\lambda x.\ (S(x) \mid \dot{d} \models \text{“}\texttt{car}\text{”}(x)^{+})(y))\ (\lambda z.\ \dot{s} \models (\texttt{shade=c})(z)^{+}))$$
$$\Rightarrow \lambda y.\mathbf{C}\times\ f_{\text{green}}(\ \lambda x.\ ((\lambda z.\ \dot{s} \models (\texttt{shade=c})(z)^{+})(x) \mid \dot{d} \models \text{“}\texttt{car}\text{”}(x)^{+})(y))$$
$$\Rightarrow \lambda y.\mathbf{C}\times\ f_{\text{green}}(\ (\lambda x.\ ((\dot{s} \models (\texttt{shade=c})(x)^{+}) \mid \dot{d} \models \text{“}\texttt{car}\text{”}(x)^{+})(y)))$$
$$\Rightarrow \lambda y.\mathbf{C} \times f_{\text{green}}(\ (\dot{s} \models (\texttt{shade=c})(y)^{+}) \mid \dot{d} \models \text{“}\texttt{car}\text{”}(y)^{+})$$

For some shade c, the result is a function from individuals to a probability value. For different values of c, this will form a distribution. However, unlike 'tall', which pulls height distributions up by some factor, 'green' will have a more constant character. Both functions will be of type $\langle p, p\rangle$, but whereas f_{tall} will have the effect of moving a distribution over heights upwards (increasing expectation of heights), f_{green} will have the effect of flattening a distribution over shades where those shades are not usually described as green, and, it will elevate the distribution for shades that are usually described as green. The function for 'green' will pull any predicate distribution it is applied to towards the same points, namely, some range of shades. This will represent why it is more reasonable to expect something described as 'green' to be some shades. This accounts for why 'green' seems to convey information on its own (about shades) in a way that 'tall' does not (about heights).

6.4.1 Connectives

For the interpretations of two declarative statements:

$$\dot{s} \models \sigma \ \mid \ \dot{d} \models \tau$$
$$\dot{s} \models \sigma \ \mid \ \dot{d} \models \rho$$

Their conjunction and disjunction will be conjunctions and disjunctions of the descriptions of the objects in the discourse situation:

$$\dot{s} \models \sigma \ \mid \ \dot{d} \models \tau \wedge_{pr} \dot{d} \models \rho$$
$$\dot{s} \models \sigma \ \mid \ \dot{d} \models \tau \vee_{pr} \dot{d} \models \rho$$

Which can be simplified using the conjunction and disjunction axioms:

$$\dot{s} \models \sigma \ \mid \ \dot{d} \models \tau \wedge_{in} \rho$$
$$\dot{s} \models \sigma \ \mid \ \dot{d} \models \tau \vee_{in} \rho$$

Values for which can be obtained using Bayes' Rule.[39] For example, in the conjunction case:

$$\dot{s} \models \sigma \mid \dot{d} \models \tau \wedge \rho = \frac{\dot{s} \models \sigma \ \ \wedge_{pr} \ \dot{d} \models \tau \wedge_{in} \rho}{\dot{d} \models \tau \wedge_{in} \rho}$$

There is, however, a non-Bayesian alternative, at least for conjunction. Since statements are interpreted as distributions, one could take the product of a distribution and then normalise the result. This approach would have the benefit of simplicity (there would be fewer correlations to learn). However, one would also lose whatever dependencies that the correlations not be included in this method reflect. I leave the final decision between these two proposals for future work.

6.5 Compositionality

As a final section to this chapter, we can now say something about compositionality. Fodor and Lepore have criticised views based on conditional probabilities for failing to capture compositionality:

> "subjective probabilities are not themselves compositional. For example, the subjective probability one assigns to the thought that brown cows are dangerous is not a function of the subjective probability one assigns to the thought that cows are dangerous, together with the subjective probability that one assigns to the thought that brown things are dangerous. If this seems not obvious, consider a world ... where there are very many things that are cows, almost none of which is dangerous, and very many things that are brown, almost none of which is dangerous, and a very small number of brown cows, almost all of which are very, very fierce. On these assumptions, the probability that something that is brown is dangerous is small, and the probability that something that is a cow is dangerous is small, but the probability that a brown cow is dangerous is as big as you please." (Fodor and LePore 1991)

[39] Bayes' Rule, stated in regular probably notation is: $P(C|A) = \dfrac{P(C \wedge A)}{P(A)}$.

To some extent Fodor and Lepore are right. Conditional probabilities taken in isolation are not compositional. The mistake is in thinking that we should take them in isolation. Put another way, individual conditional probability judgements may not compose, but (conditional) probability DISTRIBUTIONS do compose. A comparable example is 'pet fish'. The probability of something being a goldfish when described as a fish is low (it could be a cod or a tuna). The probability of something being a goldfish when described as a pet is also low (it could be a cat or a dog).

I treat 'pet' in 'pet fish' as a modifier.[40] It is not clear that there is only one range of properties that 'pet' selects (unlike shades for 'green' and heights for 'tall'). Instead, I will leave the property to be supplied by context:

pet:

$$\lambda S' \lambda P.\lambda y.\mathbf{C}\times\ f_{\text{pet}}((\ (P(y))\ (\lambda z.\ \dot{s} \models S'(z))))$$

Which could combine with a property of being some animal i.e.:

$$\lambda z.\dot{s} \models \texttt{species=a}(z)$$

The representation for 'fish' will not differ substantially from other CNs and the derivation will proceed as in previous cases.

Part of the function for pet will be assumed (over-simplistically) to provide a simple weighting over species. Domesticated species will receive an equal weighting, and non-domesticated species, an arbitrarily low equal weighting. So, attaching 'pet' to something, with respect to species, will favour domesticated ones equally, and it will reduce expectations for non-domesticated species.

With regard to species, 'fish' will be assumed to give an equal expectation for something being a cod, haddock, tuna, or goldfish. It will give an arbitrarily low reasonable expectation of something being some other species (such as a dog, a cat, or a rabbit).

In the table below, different species properties are listed in the first column. The second column gives an indication of the weighting that the function for 'pet' provides. The third column shows what expectations 'fish'

[40] Equally, however, 'pet fish' could be treated as a compound noun. In that case, the distributions for 'pet' and 'fish' could be multiplied out and normalised in a way similar to the non-Bayesian proposal for conjunction in §6.4.1. The overall effect would be similar to the way presented below. I would, however, need to give an account of how expressions of type $\langle\langle e,p\rangle,\langle e,p\rangle\rangle$ could compose to form a compound.

alone would provide with respect to species properties. Finally, the fourth column gives the weighted normalised result (the expectations for ‘pet fish’). Expectations are shown for some arbitrary object, a:

S	f_{pet} (weighting)	$p[S(a) \mid \dot{d} \models$ “fish” $(\mathtt{a})]$	$p[S(a) \mid \dot{d} \models$ “pet(fish)” $(\mathtt{a})]$
$\lambda x.\ \dot{s} \models dog(x)$	0.250	0.001	0.004
$\lambda x.\ \dot{s} \models cat(x)$	0.250	0.001	0.004
$\lambda x.\ \dot{s} \models rabbit(x)$	0.250	0.001	0.004
$\lambda x.\ \dot{s} \models goldfish(x)$	0.247	0.247	0.976
$\lambda x.\ \dot{s} \models tuna(x)$	0.001	0.250	0.004
$\lambda x.\ \dot{s} \models haddock(x)$	0.001	0.250	0.004
$\lambda x.\ \dot{s} \models cod(x)$	0.001	0.250	0.004

This gives exactly the right result. When the options are dogs, cats, rabbits, goldfish, tuna, haddock and cod, the only reasonable expectation to have, if told something is a pet fish, is for it to be a goldfish. Probability distributions are compositional relative to distributions over appropriate sets of possible outcomes.

6.6 Vagueness

On the semantics being developed, a phenomena (one that might well be called ‘vagueness’), emerges naturally from meaning representations. If terms leave us with U2 uncertainty, then sometimes there will be a clash between our intuitions of whether to apply a particular term in a particular case. U2 uncertainty (uncertainty over how to apply terms) is related to U1 uncertainty (uncertainty over what the world is like given a description). If U1 uncertainty is graded, then there will undoubtedly arise cases where U2 uncertainties clash: cases where we are no more certain about applying a term than we are about applying its negation. Such situations, I contend, are plausible explanations of so-called borderline cases.

However, we still do not have a full picture of vagueness. Recall the Austin quote we started out with in chapter 1. Explaining borderline cases alone will give a full account of vagueness. This picture will be filled out further in chapter 7 when I consider pragmatics and how the meanings of expressions interact with context.

Furthermore, some people contend that vagueness concerns truth, and more specifically, indeterminacy (chapter 1). Little has been said about truth since chapter 2, and very little by way of a positive account at all. Even if we do not end up identifying vagueness with indeterminacy, something needs to be said about truth, not just in clear cases, but also in borderline ones. Tied to this issue will be logic and which logic fits a treatment of vagueness. Truth and logic will be the topic of chapter 8.

6.7 Summary:

Learning the meaning of terms amounts to forming and building generalisations over messy data sets (chapter 5). The result of this learning process will leave a residue of uncertainty. Meanings of words do not determine how the world is (or isn't), only how the world probably is (or isn't). This view of meaning allows us to view words as carrying natural information (chapters 3 and 4). Words are useful because uses of them are correlated with states of affairs. However, on this view words do not express classical propositions. The information that words encode is neither true nor false.

Earlier in this chapter, I cited Devlin's description of situation theory as espousing a relational theory of meaning. The meaning of an expression is a relation between a discourse situation and a described situation.[41] Although the way that this idea has been developed here differs from the situation theoretic account in many ways, the basic spirit of the account remains. Individuals build connections with the world via learning (learning to decode the information carried by terms) and via communication. The reproduction of expressions for similar purposes entrenches these correlations. The relation (that is, the meaning of an expression) is a correlation. The relation is between discourse situations and described situations. People connect to this relation via semantic learning (chapter 5) and they entrench the relation via reproduction, thereby propagating the conventions that guide their use (chapter 3).

There are still several outstanding issues to cover. More must be said about vagueness and more must be said about how the meanings of terms, analysed as above, can be used for specific purposes. One such specific purpose will be to say something true, and so more must be said about what is required to count as doing so.

[41]Situation semantics also includes a reference to a speaker's connection function. I will not go into the details of these functions here.

7

Information and Truth in Pragmatics

In chapters 4 and 5, we encountered a problem with viewing language as carrying soft information grounded in correlations. The problem was, simply, that correlations are ubiquitous. In the last chapter, I suggested some ways in which this uncertainty can begin to be reduced. I suggested that generalisations and inferences drawn in the learning process can help to constrain the kinds of information that we take terms to carry. In this chapter, I will suggest that the context one is in, and the task one is engaged in can further constrain what our expectations distribute over. In the previous chapter, I argued that modifiers such as 'tall' and 'green' select-for particular types of information thus constraining the relevant information carried by nominal predicates. In this chapter, we will look at further ways in which the flow of information in communication can be refined, both by the choice of words we make, and by the circumstances in which we use them.

In §7.1, I will utilise earlier results from information theory and suggest a way to determine how much (quantitative) information a sentence carries. In §7.2, I will suggest some ways in which information can interact with the semantic information carried by expressions. In §7.3, the suggestions from previous sections will be used to give a formal analysis of one of Grice's pragmatic axioms. For the final three sections of the chapter, I will piece together the considerations of this and the last few chapters in order to generate conclusions concerning truth and vagueness.

7.1 Information

The semantics developed in the last chapter had sentences represented as probability distributions over described situations. For example, the meaning of a description of an object as red is captured as a conditional probability distribution for shades that object is, given that it is described as red. For different shades, there are different values. For some shades (the clearly red looking ones), these values will be significantly higher than all the others. There will be gradience in values between clearly red cases and clearly not red cases.[1] In chapter 4, we saw how probability values and probability distributions can be evaluated for quantities of information. Putting the semantics and the way of quantifying information together, we will be able to say, for some utterance, relative to some communicative goal, how informative it is. In this section, I will give an account of how to do this. In the next section, I will suggest how context sensitivity can enter into the account. With context sensitivity and an evaluation of informativeness, we will be able to gain a more precise understanding of some elements of pragmatics: specifically of Grice's maxim of quantity (Grice 1989a).

Recall the case of Jess and Daniel trying to choose fabric swatches (§6.3.3). Of the sixty-four swatches each has in front of them, Daniel is trying to describe one of them. Abstracting away from how context narrows down the possibilities to just one of sixty-four shades (which will have to be modelled as part of context sensitivity in the next section), we can say that there are simply sixty-four shades of colours, and Daniel's goal is to communicate his preference for one of them to Jess. This narrowing down from sixty-four to one (equivalent to six binary decisions) provides a maximum value of information for Daniels utterance (with respect to his and Jess' aims), namely 6 bits.[2]

As we did in the last chapter, imagine a special term 'green*'. Unlike 'green', 'green*' picks out any of eight shades in the swatch book (but absolutely not any of the others). Supposing that green* assigns an equal weighting to all green shades, Daniel, in using 'green*' as a classifier, would narrow the field down to one of eight swatches. Put in information theoretic

[1] No commitment is being made here to 'clear' cases being either context or interest insensitive.

[2] Recall that for a base 2 logarithm, if $x = \log n$, then $2^x = n$. In this case, there are 64 options. $\log 64 = 6$, so the situation carries 6 bits of (binary) information.

terms, Jess would still need to make three binary decisions to get down to one, so Daniel has only communicated 3 bits of information of the possible maximal 6. Put another way, Daniel's utterance using 'green*' would have an equivocation of 3 bits (3 bits of information are lost), and so he only communicates 3 bits.

'Green' isn't like 'green*'. It doesn't provide equal weightings over shades (some shades are greener than others), nor does it completely rule out shades that we might, in other circumstances, not usually describe as 'green'. A table, similar to the one from chapter 6 is below. It shows (with some different made up values) the reasonable expectations Jess should have over swatches given Daniels use of the classifier 'green':

n	$p[\dot{s} \models \texttt{like}(\texttt{Daniel})(\texttt{swatch}_n)^+ \mid \dot{d} \models \text{“Daniel likes the green swatch”}(\texttt{swatch}_n)^+]$
1	0.03
2	0.07
3	0.15
4	0.25
5	0.25
6	0.15
7	0.07
8	0.03

In the above, swatches 4 and 5 are assumed to be the most green looking. Swatches 3 and 6 are pretty clearly green (maybe a light or pale green and a dark green). Swatches 2 and 7 are perhaps only greenish. Swatches 1 and 8 are perhaps more yellowish or more blueish than greenish.

Once again ignoring the other cases for simplicity[3], the maximum amount of information that Daniel's utterance could convey will be 6 bits. Using information theory, we can evaluate how much surprisal each value would carry (given the utterance), and take a weighted sum. This will provide a measure of how much Daniel's utterance equivocates (how much information is lost, given the goal):[4]

[3]The other swatches could be assigned an arbitrarily low value and included in the table.
[4]The relevant infons are abbreviated to τ and σ.

$$
\begin{aligned}
E_{green} &= -\sum_n p(\dot{s} \models \sigma \mid \dot{d} \models \tau) \times \log p(\dot{s} \models \sigma \mid \dot{d} \models \tau) \\
&= -[2 \times 0.03 \log 0.03 + 2 \times 0.07 \log 0.07 + 2 \times 0.15 \log 0.15 + \\
&\quad 2 \times 0.25 \log 0.25] \\
&= -[-0.30 + -0.54 + -0.82 + -1] \\
&\qquad = 2.66
\end{aligned}
$$

With the equivocation subtracted from the maximum amount of information Daniel could have communicated (with respect to a choice of swatch), this means his utterance communicates $6 - 2.66 = 3.34$ bits. This reveals (perhaps unsurprisingly), that 'green' is marginally more informative than 'green*', but given that there is no English word 'green*', let's look at modifiers Daniel might use to narrow things down a bit.

Say that Daniel wants to identify swatch 6 which is darker than any of the others. Instead of just saying 'green', maybe 'dark green' would be more likely to enable him to *effectively describe* (see §6.1) a particular swatch. We could treat 'dark', much like 'tall' as a function on a probability distribution which would represent a modification of a hearers expectations towards darker shades. Spelling out such a function would, by no means, be a trivial matter, but it is in principle possible.

Using the extra modifier 'dark', we could imagine Jess' task of identifying a particular swatch might be made easier. The result on the above distribution might be as follows:

n	$p[p[\dot{s} \models \texttt{like}(\texttt{Daniel})(\texttt{swatch}_n)^+ \mid \dot{d} \models \text{"Daniel likes the dark green swatch"}(\texttt{swatch}_n)^+]$
1	0.01
2	0.01
3	0.02
4	0.03
5	0.2
6	0.5
7	0.2
8	0.03

Now it would be far more reasonable for Jess to believe that Daniel is referring to swatch 6 than any other swatch. We can again calculate the equivocation for the distribution:

$$
\begin{aligned}
E_{dark\ green} &= -\sum_{n} p(\dot{s} \models \sigma \mid \dot{d} \models \tau) \times \log p(\dot{s} \models \sigma \mid \dot{d} \models \tau) \\
&= -[2 \times 0.01 \log 0.01 + 0.02 \log 0.02 + 2 \times 0.03 \log 0.03 + \\
&\quad 2 \times 0.2 \log 0.2 + 0.5 \log 0.5] \\
&= -[-0.13 + -0.11 + -0.30 + 0.93 + -0.5] \\
&\qquad = 1.97
\end{aligned}
$$

Subtracted from 6 bits, the maximum information communicated, this gives an information value of 4.03. This is the right result, we would intuitively expect 'dark green' to be a more informative classifier than 'green'. This result is, of itself, only of minor interest. It does, however, mean that we have a way of mapping the semantic representations of utterances to an information value. No absolute significance should be attached to these values themselves. Values are calculated only for the purpose of determining if one utterance is, or would be, more or less informative than another.

In section §7.3, this method will be used to provide a formal account of Grice's maxim, paraphrased as *be no more or less informative than required for the purposes of the exchange.*

7.2 Context Sensitivity

In this section, we will look at some different ways in which context can play a role in affecting our uncertainty about the world (modelled as the probabilities assigned to situation theoretic propositions). I will suggest that context can interact with the semantics that has been provided in different ways and with different effects. Sometimes this will involve simply providing a semantic contribution not given by the words used. Other times, a further element will be introduced into the formalism to account for other knowledge, perceptions or presuppositions being made. However, suggestions made in this section will be of a far more speculative nature than the semantics of the previous chapter. At times, some suggested approaches are decidedly *ad hoc.* I hope the reader will forgive this and allow that some of these issues will be resolved in future research.

7.2.1 Problem Cases and Possible Treatments

In the last chapter (§6.2), a possible conflation was mentioned. If some classifiers (such as adjectives) are taken to be functions on probability distributions, there is a question as to what provides the probability distribution. Sometimes this seems to be nominal classifiers ('John is a tall man', 'Mary's car is green'). However, as argued by, for example, Kennedy (2007), these seem to just be special cases where a common noun specifies a comparison class. We encountered comparison classes already in chapter 1 with Delia Graff Fara's interest relative account of vague adjectives (Graff Fara 2000). Comparison classes can be explicitly asserted using 'for a' adjuncts. For example: "Jen is tall for a woman/for an adult/for a basketball player", "Kim's car is expensive for a BMW/for a radio controlled model", "Kim's BMW is expensive for a car (although not for a BMW)". Alternatively, the information in these 'for a' clauses can be provided pragmatically as a form of in-context update.

Jen is tall

As we saw in the last chapter, 'tall' is interpreted as a function on a probability distribution that is provided externally to 'tall' itself. 'Tall' merely selects height properties as the information to be considered, and shifts expectations towards greater values. The representation for 'tall' is repeated below:

tall:

$$\lambda P.\lambda y.\mathbf{C}\times\ f_{\text{tall}}(\ (P(y))\ (\lambda z.\ \dot{s} \models (\texttt{height=h})(z)^{+}))_{\langle\langle e,p\rangle,\langle e,p\rangle\rangle,\langle e,p\rangle}$$

As can be seen, we cannot merely apply this function to a type e constant (for 'Jen'), since the first argument required is of the kind of a common noun ($\langle\langle e,p\rangle,\langle e,p\rangle\rangle$). For cases like 'John is a tall man' this was provided by 'man'. In 'Jen is tall', no such lexical information is given. However, the lambda expression for 'tall' encodes a requirement for information. When this information is not provided lexically (by an expression such as 'woman'), the lambda expression for 'tall' can be understood as requiring this information to be found from the context.[5]

[5]It is possible that some of this information could come from what is known about Jen or from what is being perceived about her in context. Alternatively, with a richer account of proper names, the information that 'Jen' carries could provide this too (i.e. that she is

To access this notion of context, I will use another device found in situation semantics: *resource situations*. Situation semantics makes use of resource situations, amongst other things, for capturing the semantics of definite descriptions, indexicals, and quantification. For example, definite descriptions are challenging because they uniquely pick out an object in a local context, but frequently do not do so in a wider context. Resource situations can be used to specify a local context, relative to which a definite description is a uniquely referring term.

Suppose that it is salient that Jen, and possibly others being discussed, are basketball players. Information carried by a (resource) situation in which someone is a basketball player will look very similar to information carried by a common noun:

$$\lambda S.\lambda x.(S(x) \mid \dot{r} \models \text{basketballplayer}(x)^{+})$$

Above we have $\dot{r}$, as a resource situation parameter. We still do not want a concrete resource situation because we are interested in the kinds of correlations there are between being a basketball player and being some height. My suggestion is that this other information, if it is available or salient to the hearer, can be utilised to satisfy the requirement for information that was introduced by the utterance of 'Jen is tall'. The first part of the derivation will be similar to the last chapter yielding:

$$\lambda y.\mathbf{C} \times f_{\texttt{tall}}(\dot{s} \models \texttt{height=h}(y)^{+} \;\mid\; \dot{r} \models \text{basketballplayer}(y)^{+})$$

This is now of the right type for the constant representing Jen to be applied:

$$\lambda y.\mathbf{C} \times f_{\texttt{tall}}(\dot{s} \models \texttt{height=h}(y)^{+} \;\mid\; \dot{r} \models \text{basketballplayer}(y)^{+}) \qquad (\texttt{j})$$

$$\Rightarrow \mathbf{C} \times f_{\texttt{tall}}(\dot{s} \models \texttt{height=h}(\texttt{j})^{+} \;\mid\; \dot{r} \models \text{basketballplayer}(\texttt{j})^{+})$$

$$\Rightarrow \dot{s} \models \texttt{height=h}(\texttt{j})^{+} \;\mid\; \dot{r} \models \text{“tall”}(\text{basketballplayer}(\texttt{j}))^{+}$$

The effect here will be that one should expect the height of Jen to be in

very probably a female human being).

the upper regions of heights for basketball players. Information carried by being (or being described as) a basketball player is provided by context.

There is something incomplete about this picture. For example, I have not given an account of how resource situations are chosen to be the ones providing the information. I have also helped myself to the infon $\text{basketballplayer}(y)^+$ without saying anything about what kind of property being a basketball player is (or even could be). Notice that the same concerns do not arise for $\text{“basketball player”}(y)^+$, since the latter merely states that the expression ‘basketball player’ has been used with reference to some entity. However, this just raises the question of what the relationship between $\text{basketballplayer}(y)^+$ and $\text{“basketball player”}(y)^+$ is. These matters require further work, but will not be addressed in this thesis.

Expensive BMW

Kennedy’s example: “Kyle’s car is an expensive BMW, though it’s not expensive for a BMW. In fact it’s the least expensive model they make” is challenging for an account of (adjectival) modifiers. It suggests that, although CNs can provide comparison classes, they can be present, and nonetheless not do so. It seems that ‘Kyle’s car is an expensive BMW’ is ambiguous in meaning between *a BMW car that is expensive for a BMW* and *a BMW car that is expensive for a car.*[6]

It is not entirely clear cut as to whether we should consider this to be a purely semantic ambiguity or a syntactic one too. The first clause has the definite ring of a list about it. Just as someone might believe me to be just another verbose, prolix, philosopher, they needn’t think that I am any more wordy and motormouthed than any other philosopher. Analogously, ‘Kyle’s car is an expensive BMW’ perhaps reads more like *Kyle’s car is expensive and a BMW*. So, we have one reading in which ‘expensive’ modifies BMW and one in which it doesn’t. The first of these readings seems able to get the standard treatment where ‘BMW’ provides a comparison class as a noun and has its price distribution modified by the interpretation of ‘expensive’. However, the second is harder to capture. It seems what is needed is something like a value for expensive car:

$$\dot{s} \models \texttt{price=p(a)}^+ \;\mid\; \dot{d} \models \text{“}\texttt{expensive(car)}\text{”}(\texttt{a})^+$$

[6]However, a hearer may be able to disambiguate based on intonation.

And information about being called a BMW to combine it with:

$$\dot{d} \models \text{“}\texttt{BMW}\text{”}(\texttt{a})^{+}$$

We can then, similarly to §6.4.1, condition on this information:

$$\dot{s} \models \texttt{price=p(a}^{+}\texttt{)} \;\mid\; \dot{d} \models \text{“}\texttt{expensive(car)}\text{”}(\texttt{a})^{+} \;\wedge_{pr}\; \dot{d} \models \text{“}\texttt{BMW}\text{”}(\texttt{a})^{+}$$

Which simplifies to:

$$p(\dot{s} \models \texttt{price=p(a)} \;\wedge_{pr}\; \dot{d} \models \text{“}\texttt{expensive(car)}\text{”}(\texttt{a}) \wedge_{in} \text{“}\texttt{BMW}\text{”}(\texttt{a}))$$

In lieu of further investigation into such cases, I'll leave my analysis in this somewhat inchoate form.

7.2.2 More Resource Situations

The basic semantic shape for declarative sentences was given in the last chapter as a conditional probability statement (in fact, multiple such statements) between various described situations and a discourse situation:

$$\dot{s} \models \sigma \mid \dot{d} \models \tau$$

Take the previous example of Jess and Daniel choosing swatches. Normally, relatively independent of context, when colour terms are used, they might inform us of a very broad range of shades. In Daniel and Jess's situation, their being engaged in the activity of describing and choosing colour swatches from a swatch book restricts the number of shade candidates down to sixty-four. In yet other situations, choice may be limited even further.

One possible way to try to capture this is to condition on further information provided by resource situations. However, unlike earlier, these may now be concrete situations rather than situation parameters.[7]

[7]This would require a form of (partial) anchoring function to be reintroduced into the

$$\dot{s} \models \sigma \mid \dot{d} \models \tau \wedge_{pr} r \models \rho$$

So placed, concrete resource situations could act as constraints on, for example, the domains of objects to be considered. Additionally, they could provide information about the speaker or the hearer in the situations, or about the goals of either the conversation or the individuals engaged in it.

An example (borrowed and adapted from (Rayo 2011)) can be used to demonstrate this. Two people are going to a party. Before them are two houses and one says to the other “The party is in the blue house”. The standard situation semantic treatment of a definite description like ‘the blue house’ would be to make use of a resource situation that contains one and only one blue house. This will not do for our purposes. We cannot help ourselves to the property of being blue in a resource situation. We can only appeal to things being called blue, or being some shade or other (the shade(s) they, in fact, are). Perhaps all we need as a resource situation is the, in this case, nearby situation in which some houses are. This resource situation will then provide a localised reference class for the hearer to form reasonable expectations over. For example, simplifying a lot, following the semantic story from the previous chapter, we might start with something like the following:

$$\dot{s} \models \texttt{shade=c}_{\texttt{n}}(x) \wedge_{in} \texttt{house}(x)) \;\mid\; \dot{d} \models \text{“blue(house)”}(x)$$

In the above, x stands for some object about which the hearer has expectations relating to shades and its being a house[8], given a description of this object as a blue house. Here the speaker may be exploiting the fact that there are two houses in their locale. Let us represent, in a resource situation, information such as, that there are two houses, `a` and `b`; that the pair can see them; and what shades `a` and `b` are. Below, only the shades of `a` and `b` are represented:

$$r \models \texttt{shade=c}_1(\texttt{a}) \wedge_{in} \texttt{shade=c}_2(\texttt{b})$$

formalism.

[8]This ignores all vagueness in ‘house’.

In this example, the extra information from the resource situation is really information about possible described situations: one in which the referred-to house is $\mathtt{a}$ and so is shade c_1, the other in which it is $\mathtt{b}$ and so is shade c_2. We can, then, use the resource situation to pragmatically update the described situation and discourse situation in one of two ways. Either we can update x with object a and shade $\mathtt{c_n}$ with $\mathtt{c_1}$, or we can update x with object b and shade $\mathtt{c_n}$ with $\mathtt{c_2}$:

$$\dot{s} \models \mathtt{shade{=}c_1(a)} \wedge_{in} \mathtt{house(a)} \;\mid\; \dot{d} \models \text{“blue(house)”}\mathtt{(a)}$$

$$\dot{s} \models \mathtt{shade{=}c_2(b)} \wedge_{in} \mathtt{house(b)} \;\mid\; \dot{d} \models \text{“blue(house)”}\mathtt{(b)}$$

The interpretations of the above will be two probability values (not two distributions). One will reflect how reasonable it is to take $\mathtt{a}$ as being the referred to house. The other will reflect how reasonable it is to take $\mathtt{b}$ as being the referred-to house. The distribution for ‘blue’ will aid in determining the reference of the definite description ‘the blue house’, since when combined with information about what shades the candidate houses are, it will encode a stronger preference for objects of shades normally described as blue, than not. This means that if one of the houses is a shade often described as blue, and the other not, the probability of the house hosting the party being the former will be much higher (it will be far more reasonable for the hearer to believe the party to be in the former, not the latter).

We will stick with Rayo’s case in the next section. There we’ll look at how our formal notion of information can help characterise one of Grice’s pragmatic principles.

7.3 The Maxim of Quantity

One of Grice’s conversational maxims relates to the quantity of information carried by an utterance (Grice 1989a, p. 26):

1. Make your contribution as informative as is required (for the purposes of the exchange).

2. Do not make your contribution more informative than is required.

The purpose of this section will not be to adopt a Gricean form of pragmatics (based on implicatures arising from the violations of maxims). However, the maxims themselves are intuitive and insightful. Grice, to my knowledge, never detailed what counts as making one utterance more or less informative than another. Given the considerations of §7.1, we now have a way to model this. In the treatment of Rayo's example in the last section, we saw how resource situations can aid the resolution of a communicative goal (to identify a referent). We therefore now have a way of establishing which utterances are more goal appropriate with respect to the information that they carry.

In Rayo's case, at least part of the purpose of the exchange was to identify a house (the one with the party in it) by adequately describing that house. Rayo actually presents two distinct cases (Rayo 2011, §2):

Case 1: There are two houses before the pair. One is a blueish grey shade, the other is grey.

Case 2: There are two houses before the pair. One is the same blueish grey shade, the other is bright blue.

The cases are interesting because given the same utterance "The party is in the blue house", it seems to be more reasonable to identify the blueish grey house in Case 1 and the bright blue house in Case 2 (i.e. not the blueish grey house). Let's say that the party is in the blueish grey house in both cases. This means that, intuitively, an utterance of 'The party is in the blue house' will be effective in Case 1, but not in Case 2. By applying some information theory and what we have said about resource situations so far, this difference will be describable in terms of amounts of information and the Gricean maxim: In Case 1, an utterance of 'The party is in the blue house' is informative enough to identify the blueish grey house. In Case 2, it is not.

Call the grey shade c_0, the bluish grey shade c_1, and the bright blue shade c_2. Call the house which is grey, `a`, the house which is bluish-grey, `b`, and the house which is bright blue, `c`. Given that a description using 'blue' will correlate more strongly with things that are c_2 than c_1 and more strongly with things that are c_1 than c_0 (where this is transitive), as suggested in the last section, we get three different results:

$$p[\dot{s} \models \texttt{shade=c}_0(\texttt{a}) \wedge_{in} \texttt{house}(\texttt{a}) \mid \dot{d} \models \text{“blue(house)”}(\texttt{a})\ ] \quad = \text{Low}$$

$$p[\dot{s} \models \texttt{shade=c}_1(\texttt{b}) \wedge_{in} \texttt{house}(\texttt{b}) \mid \dot{d} \models \text{“blue(house)”}(\texttt{b})\ ] \quad = \text{Higher}$$

$$p[\dot{s} \models \texttt{shade=c}_2(\texttt{c}) \wedge_{in} \texttt{house}(\texttt{c}) \mid \dot{d} \models \text{“blue(house)”}(\texttt{c})\ ] \quad = \text{Even Higher}$$

This goes some way to account for why an utterance of 'The party's in the blue house' is successful in identifying/picking out b in Case 1 but not in Case 2. Given a choice between low and higher values, opt for higher. Given a choice between higher and even higher, opt for even higher.

Information comes in when we look at alternative utterances that could be used to achieve the goal of describing the blueish grey house in Case 2. One possibility would be to use a more specific colour term such as 'bluey-grey'. That might give higher results for b than for c. As we saw in §7.1, more specific colour terms carry more information (think of Daniel using 'dark green' instead of just 'green'). Hence we can see how the first part of Grice's maxim applies: If 'blue' won't do for the purposes of the exchange, use something more informative, say, 'blue-grey'. We can also see how the second part operates: If the less informative 'blue' will do, don't use the overly informative 'blue-grey'.

7.4 Truth

In chapter 2, I argued that truth conditions cannot form the basis for an account of meaning. This led to the development of an account of meaning in terms of information, and what it is reasonable to believe, given a linguistic signal. However, as I have remarked, when the meanings of terms are characterised in this way, questions of truth are simply not addressed. For example, when John is described as tall, I have argued that, in context, relative to some information (comparison class), there are expectations that we can reasonably have with respect to his height (however that is encoded). I have represented these expectations using probabilities, as probability distributions over situations in which John is some height.

Now, for lower heights, John being that height will have a low value. For higher heights (up to a point), John being that hight will have a higher value.

And yet, John will actually be some height. Suppose John is actually one of the lower heights. Should the WORDS and what they mean be called false? Likewise, suppose John is actually one of the higher heights. Should the WORDS and what they mean be called true? These may be valid questions, but what needs to be made clear is that the meanings of the words and the information they carry does not answer them. The reason for this is simply that if the presence of some signal is correlated with certain states of affairs, then even if it turns out that a less probable state of affairs obtains, the signal is not inconsistent with this. The signal is not thereby false. Furthermore, if a more probable state of affairs obtains, the signal is not thereby true. Indeed, that different states of affairs obtain, with varying frequencies (i.e. that they are correlated to various degrees with the presence of the signal), is exactly what the signal does represent.

And yet, it is not as if there isn't something wrong, frequently, about calling someone of small stature 'tall'. However, it is not the words that are wrong in this sense but the use that they have been put to. If I want to point John out to you at a party, I am likely to fail if John is of smaller stature than most others present and I describe him using the word 'tall'. If Jess wants to paint her walls with teal paint, she oughtn't to ask to see the selection of yellows that the paint supplier has in stock.

We might ask, "If the meanings of words and sentences (and the way the world is) do not DETERMINE whether they are true, what does?" But to do so might risk asking a question with no answer. Why should we expect there always to be a DETERMINER of truth value? To do so would be to expect, for some sentence, there to be a set of necessary and sufficient conditions such that if the world matches them, the words are made true. Those necessary and sufficient conditions would be necessary and sufficient conditions for truth (truth conditions) and it is precisely such conditions that I have argued cannot be the sorts of things words carry on their own.

People speak the truth (and also fail to do so). It is true that, currently, London is the city I live in. That is, for many purposes my using the sentence 'I live in London' will be true. Whether or not I am speaking the truth can, on some particular occasion, turn on non-linguistic factors. For example, sometimes it will be important not to use 'London' to describe where I live, if, in that context, 'London' would be taken to refer to The City of London (in which, I have never resided). Yet what determines how this will probably be taken may be most affected by the goals, aims and purposes of those with whom I am in conversation.[9]

If an individual has good command of a language, they will usually form expectations in line with how words are used in that language. They will also be able to use words to affect the expectations of others.[10] Sometimes, either through ignorance or malice, people use words in such a way that others will come to expect that things are the way that they are not. Given the information the words used carry (and probably some other factors about the situation), it would be reasonable for them to think that the world is some way that it is not. On other occasions, either through accident or purpose, people use words which give their interlocutors exactly the right kinds of expectations as to how the world is.

At least for the kinds of simple terms that have been the main focus of discussion in all of the above, the two kinds of cases seem like good descriptions of when words are used to say something false (as in the former case) and true (as in the latter).

I will now discuss some possible implications of this view in a very brief and sketchy manner. Each of the issues touched upon could be the subject for a separate research project.

Pragmatics and Truth

I take this kind of view to be akin to the pragmatic view of truth in Travis:

> "Pragmatics [...] is the study of properties of words which depend on their having been spoken, or reacted to, in a certain way, or in certain conditions, or in the way, or conditions, they were." (Travis 2008d, p. 109)

In this sense of pragmatics, truth is a pragmatic property of words (one that words may sometimes have or sometimes lack dependent on practical factors such as the ones listed in the quote above). If conditions both in the situation of utterance and the described situation can change what it is reasonable to expect the world to be like (given what has been said), and if our reasonable expectations guide our judgements about truth, then the conditions in situations of utterance and situations of description can effect

[9]This idea is in contrast to Lasersohn (1999) who thinks that most cases of what he calls 'loose speaking' are literally false. On the view defended in this thesis, one could say, for example, that Mary arrived at three o'clock (when she arrived at one minute past) and be speaking the truth. On Lasersohn's view, one would, strictly speaking, be saying something false, although close enough to the truth to be acceptable.

[10]This needn't involve reasoning about the expectations of others, however.

whether or not the same words used on different occasions count as true or false (even if the object described and the properties it has remain constant).

The view is also related to Austin's speech act correspondence view discussed in §2.1.2. On Austin's view, (someone's) words are true if the situation they describe is of a type with the kind of situation they conventionally describe. Whether the actual situation is of a type with the kind of situation that those words conventionally describe may itself, however, require an application of reason. Some of those reasons may depend on what is being done with those words at the time.

Objectivity

The information that terms carry is based on how they are used. There is, then, a sense in which, whatever degrees of belief one has about the world being some way (given what has been said), one can be wrong: one's expectations can be unreasonable if they differ from the information carried by the signals received. Sources of error are no doubt ubiquitous and affect (infect) us all, but we seem to get by a lot of the time. Not all, or perhaps even not many, uses of language rest on the details being precise.

Even if we accept truth as being a pragmatic property of words used on some occasion, this doesn't remove all objectivity. The information that is carried by words is such that how an individual uses them will be, statistically, pretty much irrelevant. So what it is reasonable to expect given the use of some words will also retain an aspect of objectivity despite any individual's errant use.[11] Furthermore, for many cases at least, whether things are as is reasonable for us to expect them to be, will not be a controversial issue. Truth needn't always be a matter for debate, and there needn't be room made for anything like true-for-me or true-for-you.

Science

We should not take a pragmatic view of truth to be relativistic or anti-scientific.[12] The view on truth described above should be seen as being firmly in the scientific pragmatist tradition of C. S. Peirce. Take the following quote from Peirce (emphasis is my own):

> "There are real things, whose characters are entirely independent

[11] However, if everybody started using a term in a different way, the information it carries would start to shift.

[12] If there is relativism in this view, it would only be of a Quinean ontological sort (Quine 1968).

> of our opinions about them; those realities affect our senses according to regular laws, and, though our sensations are as different as our relations to the objects, yet, by taking advantage of the laws of perception, we can ascertain BY REASONING how things really are, and any man, if he have sufficient experience and reason enough about it, will be led to the one true conclusion" (Peirce 1877, p. 12)

The above passage speaks of reality, but importantly, of the discovery of reality by both perceiving the world and using our reason. In fact, Peirce was, in that passage, informally defining the scientific method of fixing belief.

One could read Peirce as preempting some of the ideas from Bayesian learning models and Cognitive Science that were mentioned in chapter 5.[13] The idea behind such models is to account for learning (both in terms of scientific learning about the world and the kind of learning we do as children) as an integrated process of perception and reasoning; of inference drawing and hypothesis testing. When I talk about words making certain states of affairs more reasonable to believe. This can be read as wholly consistent with this broader and wholly scientific approach to study.

By repetition (of experiments/of learning experiences), and in science, by reproducing results, uncertainty is reduced to acceptable standards. In science, acceptable standards may be such that it would be absurd to doubt the results. Practically speaking doubt is eliminated. Actually speaking, there might be a minute chance of some kind of error. In everyday speech, we are not usually so careful, but reasoning applies there too. Our everyday terms are used to apply to many domains and this flexibility can mean that we must reason using information external to that carried by the terms if we want to determine whether those words were used to say something true.[14]

Once meaning and communication are freed from truth (or perhaps, once truth is freed from the burden of having to support meaning and communication), truth becomes a subtle and complex phenomena. Wilde was getting at something in *The Importance of Being Earnest* when he wrote "The truth is rarely pure and never simple".

[13] Peirce introduces his monograph by extolling the virtues of applying "statistical methods" to new areas of inquiry. He devotes the entire third part of this classic work on pragmatism to probabilistic reasoning.

[14] An excellent analysis of how pragmatic ideas about truth enter into scientific terms and conversations can be found in (Davies 2011, ch 3).

7.5 Travis Cases

Part of the support for disassociating truth and truth conditions from meaning rested, in chapter 2 on an example from (Travis 2008d). Pia and her painted leaves is just one example of what, in philosophy of language circles (at least in London), are known as *Travis cases*. The Pia case was deployed by Travis to argue that the sentence "The leaves are green" cannot have a set of truth conditions as part of its meaning. But now that something of a pragmatic story has been given, I will try to give some analysis of what is happening in a case of this kind.

Here, again, is the case as it originally appears:

> "A story. Pia's Japanese maple is full of russet leaves. Believing that green is the colour of leaves, she paints them. Returning, she reports, 'That's better. The leaves are green now.' She speaks truth. A botanist friend then phones, seeking green leaves for a study of green leaf chemistry. 'The leaves (on my tree) are green' Pia says. 'You can have those.' But now Pia speaks falsehood." (Travis 2008d, p. 111)

Alternatives can be found throughout Travis's work. For example:

> "Consider the sentence 'The ball is round', and two cases of its use. *Case A:* What shape do squash balls assume on rebound? Pia hits a decent stroke; Jones watches. 'The ball is round' she says at the crucial moment. Wrong. It has deformed into an ovoid. Jones did not say the ball to be as it was, so spoke falsely. *Case B:* Fiona has never seen squash played. From her present vantage point the ball seems a constant blur. 'What shape is that ball?', she asks. 'The ball is round', Alf replies; truly, since it is the sort of ball a squash ball (and this one) is. It is not, e.g., like a very small rugby ball." (Travis 2008b, p 97)

This latter example is reminiscent of a point Austin makes in *Truth*:

> "We say, for example, that a certain statement is exaggerated or vague or bald, a description somewhat rough or misleading or not very good, an account rather too general or too concise. In cases like these it is pointless to insist in simple terms whether the statement is 'true or false'. Is it true or false that Belfast is north of London? That the galaxy is the shape of a fried egg? That

> Beethoven was a drunkard? That Wellington won the battle of Waterloo? There are various degrees and dimensions of success in making statements: the statements fit the facts always more or less loosely, in different ways on different occasions for different intents and purposes. What may score full marks in a general knowledge test may in some circumstances get a gamma." (Austin 1950/1979, pp. 129-30)

These three passages each strongly suggest that things like circumstances, goals, and purposes can affect our judgements regarding the truth values of statements. However, in a very direct sense, because there is no truth value *simpliciter* for the sentences used to make these statements, circumstances, goals and purposes can also affect truth values of the statements themselves.

To capture this sort of interaction is by no means an easy task and I have little idea of how to begin to systematically go about doing so. However, in the semantics developed so far, we have used resource situations to incorporate extra information into described and discourse situations. For example, that a scientist requires green leaves for a green leaf science experiment may lead us, reasonably and rationally, to expect her to be looking for leaves that are green even without paint. If a pilot flying from London to Belfast asks for a compass bearing to navigate by, being told that London is north of Belfast would be hopeless. Perhaps this information (about what green leaf scientists probably require of their test samples, and what pilots typically need when trying to navigate) can be placed into the picture through resource situations, as something like background knowledge that is being utilised in a situation. However, it is another question how we manage to focus on the right information for the right purposes.

It seems likely that we have all had to learn to incorporate the effects of purposes and goals (and employ our background knowledge) into what we say and do. We must learn this as much as we have had to learn linguistic concepts. Yet, even though the complex and puzzling Travis case data and the equally thought provoking examples from Austin were problematic on a truth conditional conception of meaning, they are not disastrous for the view of meaning as correlative, statistical information. On the truth functional view, all our knowledge of the world, the know-how of applying it, and the sensitivity of truth values to purposes and goals, was built into the meanings of terms themselves. In chapter 2, our conclusion was that this would make meanings impossible to learn, and, even if learnable, impossible to communicate with. On the correlation view, such know-how and sensitivity to purposes and goals is placed firmly outside of the meanings of terms and instead is held to be something that must be learnt in itself. This splitting up of the learning task is, again, in line with Bayesian learning models in cognitive science. Those models (such as in (Kemp and Tenenbaum 2008)) employ domain general learning strategies to infer domain specific constraints. In the above, that is arguably comparable to using general abilities to learn a structure in which other linguistic information can be situated and integrated.

I will try, very tentatively, to elaborate how the semantics and pragmatics developed could analyse one of the above examples. Our experience of navigation situations will correlate more strongly with situations in which agents require information about more specific directions than it does with situations in which agents require information about more general latitudinal comparisons (in the sense of, for example, everything in the Northern Hemisphere being north of everything in the Southern Hemisphere). Learning to navigate and learning to help others to navigate will involve becoming sensitive to such differences in the strengths of these correlations. This does not imply that directions of the more general sort will never be needed in navigation situations; merely that there is a lower probability that they will be. This learning process can be deployed in a navigation situation when asked by a pilot which direction to fly to get from London to Belfast. We will assign a much lower value to her requiring a general direction, and a much higher one for a specific (set of) direction(s).

Put in the formal sense:

$$p[\ \dot{r} \models (\texttt{Needs specific directions})(a) \mid \dot{r}' \models (\texttt{Needs to navigate})(a)\ ] = \text{high}$$

$$p[\ \dot{r} \models (\texttt{Needs general directions})(a) \mid \dot{r}' \models (\texttt{Needs to navigate})(a)\ ] = \text{low}$$

This learnt (probabilistic) knowledge can be deployed in a navigation situation. Doing so will mean that ‘Belfast is north of London’ will get a lower value, whereas ‘The direction from Belfast to London is n degrees’[15] will get a higher probability. This will reflect why it seems more reasonable to say, in a navigation situation, that the former is false and the latter is true. The

[15]For whatever the actual value is.

former would be false because in uttering ‘Belfast is north of London’ in a navigation situation, one will be interpreted as saying something more like ‘Belfast is heading 000 degrees from London’. To utter the former to the pilot in the navigation situation, whilst intending to communicate the more general thought would be misleading, and it would be an incorrect use of words, relative to the purpose of navigating.

7.6 Ways to be Vague

This thesis began with a quote from Austin, reprinted below:

> "‘Vague’ itself is vague. Suppose I say that something, for instance somebody’s description of a house, is vague; there is a quite large number of possible features—not necessarily defects, that depends on what is wanted—any or all of which the description might have and which might lead me to pronounce it vague. It might be (a) a rough description, conveying only a ‘rough idea’ of the thing to be described; or (b) ambiguous at certain points, so that the description would fit, might be taken to mean, either this or that; or (c) imprecise, not precisely specifying the features of the thing described; or (d) not very detailed; or (e) couched in general terms that would cover a lot of rather different cases; or (f) not very accurate; or perhaps also (g) not very full or complete... there is not just one way of being vague, or one way of not being vague, viz. being precise." (Austin 1962, pp. 125-6)

With the analysis of the meanings of terms that has been given, and with a pragmatic approach to truth, different elements of the story told so far can come together to characterise the ways that terms can be used to be vague in many of the above senses. In §7.1, we saw that the use of intensifiers and further terms (such as ‘dark’ and ‘light’ for shades) can increase the amount of information conveyed (can make the expectations a hearer could reasonably have more specific and less wide ranging). We also saw how the goal of an exchange can place different demands on how much information needs to be conveyed. One analysis of many of the senses above (namely, (a), (c), (d), (e), (f), and (g)) could be put in terms of failing to provide enough information for some goal. The case of the pilot above is a good example: The rough/unclear/imprecise description, ‘Belfast is north of London’ would be vague and unhelpful precisely because it does not carry the information the pilot needs.

Sense (b) seems to be a bit different. Little has been said so far about ambiguity (save from a discussion of ambiguity* in §1.6). But ambiguity, too, could be given a similar analysis in some cases. Where the same word carries two or more distinct sets of information, sometimes we will need to add more information to what we say to ensure that one set rather than another is used by our interlocutor.

7.7 Comparison With The Literature

7.7.1 MacFarlane's Fuzzy Epistemicism

MacFarlane (2009b) criticises the assumption that degree theories should treat the values for degrees as artefacts of a formalism. He defends the view that we can have hidden unknown or unknowable semantic facts, uncertainty over where what these facts are, and degrees of truth. He calls this position *Fuzzy Epistemicism*. What is unknown on MacFarlane's model is what exact degrees of truth a statement has given the state of the world.

MacFarlane is interested in accounting for mixed cases of vagueness and uncertainty. He argues that reasoning about vagueness is best modelled with Fuzzy Logic (I will criticise the reasons for thinking this in chapter 8), but that reasoning in conditions of uncertainty is best modelled with classical probability calculus. However, given that fuzzy degrees of truth are as numerous as the real numbers in the range [0,1], and, given that MacFarlane is committed to there being absolute values for propositions (not mere artifacts of a model), he is an epistemicist about what exact value any proposition has.

Whereas on my view, the meanings of statements are given as distributions over described situations, on MacFarlane's view, (our attitudes towards) statements are represented as distributions over degrees of truth. I will say more on degrees of truth in chapter 8, but the connection between my and MacFarlane's account is strongest in how he calculates such distributions.

Call the proposition that Tom is bald, T, and the proposition that Tom has n hairs, H_n. The distribution for having a degree of truth x is given by the following ($[\cdot]$ denotes the fuzzy interpretation function):

$$Pr([T]=x) = \sum_{0\leqslant n<500} Pr(H_n) \times Pr([T]=x|H_n)$$

Translating MacFarlane's work into my own vernacular, this models how certain one should be of a baldness judgement, given the correlation between baldness judgements and hairiness, weighted against the prior probabilities of how hairy Tom is.

MacFarlane's degrees of truth can, perhaps, be related to probabilities of judgements, given a situation. In which case, his apparatus could be used to indicate higher-order probabilities of the kind I have been discussing. For example, the probability of a degree of truth could be taken to be the probability of a probability that a baldness judgement would be made, given uncertainty over how hairy someone is.

So understood, MacFarlane and I differ with respect to what we are trying to explain. I am interested in uncertainty about states of affairs, given that someone has made a verbal judgement. He is interested in uncertainty that remains with respect to the truth-values of judgements. Where we fundamentally differ is what logic we think should govern borderline cases and in how we understand degrees of truth. Both of these matters will be discussed in chapter 8.

7.7.2 Lassiter, and Frazee and Beaver's Epistemic Models

Although Lassiter (2011) and Frazee and Beaver (2010) describe their positions from different perspectives, they amount to similar approaches. Lassiter's article, though marginally later was developed independently and is fuller in detail. I will focus on it.

Similarly to my approach, Lassiter is interested in uncertainty in communication and worried about sharp boundaries. He is less worried about truth-conditional semantics in general, however. Rather than implementing probabilities directly into the meaning representations of words, Lassiter incorporates metalinguistic uncertainty over what is being expressed (in terms of a sharp proposition). He also includes uncertainty about the world, but we'll focus on the former case.

Lassiter inherits, from other work in the scalar adjectives literature, the idea that their semantics should incorporate thresholds. In simple terms, what Lassiter models is uncertainty over where the threshold is when a term like 'tall' is used. We can, relative to a context, be pretty certain that thresholds are not too high or too low. Lassiter models this as a distribution over precise model theoretic objects which have sharp cut-offs. The effect is that if we learn someone's height, we can get a value that reflects the probability that that individual is tall.

However, Lassiter's emphasis is not on the sharp boundaries that such a position implies. For him (and for Frazee and Beaver), communication using 'tall' is about approximating, roughly, the standard in play for what would count as tall. It is these standards that we are uncertain about when someone uses the expression 'tall'. However, this general picture has the effect of implying that there is, nonetheless, a standard for 'tall' in every context. I have argued that approximation of standards is not what is encoded by our words. This amounts to an argument that uncertainty should enter at the level of describing the model theoretic object, not at the level of evaluating which classical model theoretic object is in play.

The difference between our positions could, however, amount to one of levels of idealisation. Lassiter's account is a big improvement on the standard classical model and it does address concerns about communication. For example, on Lassiter's view, others' uses of terms can be used an evidence to infer what they mean. However, the fact remains that the motivation for appealing to precise languages must still be given. If it turns out that we rarely, if ever, coordinate sufficiently to settle on exactly where a threshold for a term for 'tall' should be, then why have such thresholds written so centrally into the semantics of such terms? We can drop precise languages and have a far more direct connection to what our terms mean by adopting the picture I have proposed. This gives us a way of linking experience of others' linguistic behaviour directly to an account of semantic learning.

7.7.3 Cooper et al.'s Probabilistic Type-Theoretic Approach

Semantic learning is at the forefront of (Cooper et al. 2013), which is also the closest position to my own.[16] I strongly suspect that the core of the two positions will be pretty inter-definable, although, again, there are differences of emphasis.

In chapter 6, I mentioned two kinds of uncertainty. U1 uncertainty was uncertainty about the world, given a description of it. U1 uncertainty is, effectively, what my account describes. Arguably, Cooper *et al.* focus on U2 uncertainty. U2 uncertainty is uncertainty about how to use words, given some known or perceived way the world is. I will say more about Cooper *et al.*'s formalism in a moment, but it is first worth remarking that, if modelled as I have presented, U1 and U2 uncertainty are inter-definable. The main weight of the meanings of utterances in my account rests on

[16]I only read a draft of this paper in the closing months of writing this thesis.

conditional probabilities of the form $p[\dot{s} \vDash \sigma | \dot{d} \vDash \tau]$. In other words, the probability of the world being some way, given a description of it. If, indeed, Cooper *et al.*'s account describes U2 uncertainty, then it is possible that their results could be simulated via the use of the alternate conditional probability: $p[\dot{d} \vDash \tau | \dot{s} \vDash \sigma]$ (or the probability that some description will be used, given that the world is some way). Importantly, given the priors $p[\dot{s} \vDash \sigma]$ and $p[\dot{d} \vDash \tau]$, these two conditional probabilities are simply ratios of each other.

Cooper *et al.*'s semantics uses a rich theory of types. The simple type theory used in chapter 6 provided domains for basic types. Complex types were then constructed as functions of basic types. In rich type theories, types should not be thought of in these extensional terms, but instead as something like useful ways of classifying things (Cooper 2012, p. 275). However, propositions are also types. For example, the proposition *David runs* can be seen as a situation type (one in which David runs), and is true iff there is a situation in the world in which David runs. In Cooper *et al.*, agents are modelled as making probabilistic judgements about classifications. Propositions are still types, but judgements reflect the probabilities that there is a situation of the right type.

Of particular interest is how these type judgements are grounded. Central to Cooper *et al.*'s account is semantic learning. On their learning model, a learner is exposed to multiple situations. In each one they make a judgement about whether the object they are shown is an apple or not an apple. After each judgement, the oracle that models the adult speaker gives a judgement. Initially, the learner cannot make a judgement, but after a few exposures to adult judgements, has enough data to begin to make judgements themselves. In Cooper *et al.*'s model, 'apple' judgements are based on four properties (two colours and two shapes). Simplifying a little, when faced with a new object to classify, the value of the judgement is calculated as the conditional probability $p(apple|observed\ properties)$.[17] To calculate that conditional probability, the learner uses priors and conditional probabilities such as $p(property|apple)$, both of which are estimated directly from the adult judgements they have witnessed.

Translating more into my own vernacular, one could see probabilities of 'apple' judgements as approximating probabilities of discourse situations, and we could see probabilities of properties as approximating probabilities of described situations. The two stages of Cooper *et*

[17] Where the output judgement is decided by whether an 'apple' or an 'not-apple' judgement receives a higher value.

al.'s learning account can then be described as: (i) Estimating probabilities of the form $p[\dot{s} \models colour(x)^+]$, $p[\dot{s} \models shape(x)^+]$, $p[\dot{d} \models \text{“}apple\text{”}(x)^+]$, $p[\dot{s} \models colour(x)^+|\dot{d} \models \text{“}apple\text{”}(x)^+]$, and $p[\dot{s} \models shape(x)^+|\dot{d} \models \text{“}apple\text{”}(x)^+]$ directly from witnessing adult speakers' linguistic behaviour. (ii) Using those values to calculate, of a novel object/context, the probability $p[\dot{d} \models \text{“}apple\text{”}(y)^+|\dot{s} \models colour(y)^+ \wedge shape(y)^+]$.

In chapter 5, I described, albeit informally, only the process in (i). Cooper *et al.* show with precision how the kind of information that can be learnt from adult speakers' linguistic behaviour can be turned into a classifier judgement. These judgements will not always be probability 1. The uncertainty that < 1 judgements reflect, is, I would argue, just what I described as U2 uncertainty. A closer examination of the two accounts should be made once both have been finalised. It would certainly be useful if I were simply able to adopt Cooper *et al.*'s learning model.

One distinct contribution the account in this thesis makes is in the treatment of adjectives. Cooper *et al.*'s learning account shows how a nominal like 'apple' could be learnt from basic colour and shape observations. I have argued that colour concepts (as well as other vague concepts) should be viewed as functions on distributions given by nominal classifiers.

To end this chapter, I will demonstrate how my own account could be further understood in Cooper *et al.*'s terms. In the second column below is the toy distribution we had for 'John is a tall man' in chapter 6. With the assumption that there are now four possible discourse situations (“tall man”, “not-tall man”, “tall, but not a man”, “neither tall nor a man”), the prior for the discourse situation is 0.25. Via a simple application of Bayes' rule, values can then be computed for the probabilities of “tall man” judgements, given a height for that individual. This is shown in the third column:

h	$p[\dot{s} \models \texttt{height} = \texttt{h}(\texttt{j})^+ \mid \dot{d} \models \text{“tall man”}(\texttt{j})^+]$	$p[\dot{d} \models \text{“tall man”}(\texttt{j})^+ \mid \dot{s} \models \texttt{height} = \texttt{h}(\texttt{j})^+]$
$h < 150$	0.01	0.03
$150 \leqslant h < 155$	0.01	0.03
$155 \leqslant h < 160$	0.01	0.03
$160 \leqslant h < 165$	0.01	0.03
$165 \leqslant h < 170$	0.01	0.03
$170 \leqslant h < 175$	0.02	0.06
$175 \leqslant h < 180$	0.06	0.17
$180 \leqslant h < 185$	0.11	0.30
$185 \leqslant h < 190$	0.22	0.61
$190 \leqslant h < 195$	0.32	0.88
$h > 195$	0.22	0.61

The values in the final column, I suggest, are equivalent, at least in spirit, to Cooper *et al.*'s type judgements.

8

Avoiding a Degree of Trouble

In the last two chapters, a formal model for the meanings of vague terms was given (based on a statistical notion of information), and some suggestions were made for how context and further information can affect what is communicated in using those terms. Furthermore, it was argued that, on such a view, truth is a pragmatic concept, at least insofar as, given some described situation, the truth-conditions of an utterance are not determined by the meanings of the words alone.

In §8.1, I will argue that indeterminacy arising from vagueness should be understood as insoluble metalinguistic uncertainty. This account of indeterminacy allows for an alternative interpretation of Edgington's verities as records of mounting (metalinguistic) uncertainty. In a limited domain, these uncertainties can also be understood as degrees of truth. Both of these ideas will be described, but this chapter will mostly consider the latter. In §8.2, Bayesian degrees of truth will be defended against arguments found in the vagueness literature. The latter notion will be discussed further in §8.3 and in chapter 9. In §8.3, I will evaluate the impact of accepting Bayesian degrees of truth on the principle of bivalence and examples of disquotation.

8.1 Verities and Indeterminacy

In §6.6, I gave a characterisation of how borderline cases can be analysed. Since our semantic representations are graded, and since our classification judgements are related to these semantic representations, we could easily find ourselves in a situation where, given that our words mean what they do, we would have as much reason to classify an object as, say, 'red' as to classify it as 'not-red'. Nothing in the meaning of 'red' will decide the case for us. In this way, vagueness *qua* borderline cases emerges naturally as a byproduct of two phenomena/processes: (i) the messy correlations that exist between uses of words and ways the world is. (ii) our learning and representing these correlations (semantic learning).

In some borderline cases, the circumstances we are in, or the aims and goals we have may push us in one direction or the other (chapter 7). What might, in some other circumstances, have left us torn over classification, may, in our current circumstances, be clearer. Lucy's hair might be a borderline case of brown when describing it in some circumstance, but not necessarily when she is the only person present who doesn't have bleached hair. This notwithstanding, we still want to be able to say something about truth when we find ourselves thoroughly torn. In some circumstances, for some purposes, there may be no information available (be it known or unknown) that would help tip the balance of reasons in favour of one classification or another.[1]

We still want to be able to detail what sorts of rules of inference govern reasoning in borderline cases. For example, we want to be able to say how conjunctions or negations etc. interact with failures to come down on one side of the true/false division. For example, should conjunctions of borderline cases push us closer towards judgements of truth than judgements of falsity?

In chapter 1 (§1.7.2), when discussing pure degree theories of vagueness, Edgington's Bayesian degree logic (BL) looked promising (reprinted below):

$$\begin{array}{lrcl} (\neg_{BL}) & v(\neg\alpha) & = & 1 - v(\alpha) \\ (\wedge_{BL}) & v(\alpha \wedge \beta) & = & v(\alpha) \times v(\beta \mid \alpha) \\ (\vee_{BL}) & v(\alpha \vee \beta) & = & (v(\alpha) + v(\beta)) - (v(\alpha \wedge \beta)) \end{array}$$

In §1.7.2, three possible interpretations for verities were discussed. In

[1] Sorites paradoxes are designed to bring such circumstances about. The paradox will be addressed in the next chapter.

this chapter, I will outline a defence of the epistemic understanding of verities (node [00] on the tree in §1.3).

I will not be arguing in favour of a simple degrees of truth view (node [01]). Reading verities as degrees of truth (where Bayesian logic simply replaces fuzzy logic), leaves it mysterious as to what degrees of truth are, and why they should be modelled with a probability calculus. The epistemic reading will prove to be less mysterious.

Edgington argues against an epistemic interpretation of verities. I agree. Not every case of uncertainty is a case of indeterminacy. However, below, I outline how to understand indeterminacy as a species of metalinguistic uncertainty (as opposed to uncertainty about the world).

8.1.1 Indeterminacy

The main idea I defend here is that there is a connection between reasonable belief and indeterminacy. I argue that, in special cases, we can be left with irresolvable metalinguistic uncertainty as a result of the kind of information that terms carry. In such cases, there is no reason to favour one truth value over another. Such special cases, I suggest, are cases of indeterminacy. Furthermore, if indeterminacy is a form of metalinguistic uncertainty, we will have a reason why the logic of such indeterminacy should be Bayesian in flavour.

Truth, on the pragmatic view, is not determined by meaning, rather it is informed by it. Truth is a property that some statement may have, and whether or not it has it depends how close the reality of the described situation is to what one would expect, given the information that utterances carry. However, how close the match needs to be is affected by what the interests and purposes of those involved in an exchange are.

A description may lead us to expect some situation to be of some type (with some weighting). The negation of that description may also lead us to expect, with the same weighting, a situation of an incompatible type. For example, for some height $\mathtt{h}$, we may have low expectations that some man is $\mathtt{h}$cm in height, given a description of him as tall. But we might also have equally low expectations of him being $\mathtt{h}$cm in height, given a description of him as not tall. So, should we come across someone of this height, we would have no more reason (given the meaning of 'tall') to describe him as tall than we do for describing him as not-tall.

Should this case arise, we would be in the perplexing situation of no longer having reasons for an utterance of either, 'He is tall', or of 'He isn't tall', being true (or false). Furthermore, in some situations, there may simply be no reasons to be found. Words can, literally, fail us. This happens when there is no information, either that we have, OR THAT THERE IS TO HAVE, for classifying things one way or the other. This is a form of uncertainty, and one in which the uncertainty is irresolvable. However, unlike the uncertainty of epistemicism, which concerns uncertainty about how things ARE (one way or another), at issue here is metalinguistic uncertainty: uncertainty over how to behave linguistically.

To emphasise, this is not anti-realism of the sort raised by Dummett (1959). If all the facts were in, we would know if there were an odd or an even number of stars in the universe.[2] When all the facts are in, the information our words carry can nonetheless fail to provide sufficient reason for a linguistic classification.

Metalinguistic uncertainty is not so much about knowledge as it is about reasons for action: which way to classify something. I contend that metalinguistic uncertainty is what we call 'indeterminacy'.[3] Here, we have a special situation in which the meaning of the classifier, whether we know it or not, gives us no reason to make a classification one way or the other.[4] Put another way, what terms mean (and how this interacts with circumstances), can leave us with conflicting expectations, but also with no way out of the conflict. When stuck in this way, neither a classification nor its negation has a privileged position as the true classifier. When stuck in this way, it is indeterminate which classifier applies.

8.1.2 Edgington's Examples

To what extent does this count as collapsing indeterminacy into uncertainty? This question is important, given an argument of Dorothy Edgington's (Edgington 1997). Edgington points out that uncertainty can play a different conceptual role for us than indeterminacy. Where uncertainty is credence, we can be as certain as not that ϕ. This guides our behaviour in a way that assigning a verity of 0.5 to ϕ does not. Here is one of her examples:

> "I want a cup of coffee; tea would do as second best. My cupboard is bare. I call on a colleague. C will almost certainly give me

[2]Unless borderline (proto)stars created metalinguistic uncertainty.

[3]At least of the sort often referred to as 'semantic indecision'. Other kinds of indeterminacy, such as reference failure, might have another source.

[4]Even better, we have 'no sufficient reason'. When we talk of reasons for action, vagueness is situated there too. This will be discussed in the next chapter.

> coffee; it's about 50-50 whether D will give me coffee or tea; and E will almost certainly only have tea." (Edgington 1997, p. 313)

In this case, given the preference for coffee, the rational ranking of whose office to go to is C, then D, then E. Edgington contrasts the case with one involving indeterminacy:

> "Now let's change the story about D. D, the archetypal absent minded genius, has curious ways of making hot drinks. He is liable to ignore the difference between tea pots and coffee pots [...] You get a cup of something from D, and it is genuinely indeterminate whether the word 'coffee' can be applied or the word 'tea', it is a borderline case." (Edgington 1997, p. 313)

As Edgington points out, the indeterminacy of the second case does not have the same impact on our rational ranking of whose office to go to as the uncertainty of the first case. Whether or not one prefers ending up with borderline tea-coffee to both, either, or neither, coffee or tea is not, in any way, determined by a preference for coffee to tea. Given a preference for coffee over tea, I might, rationally, love, detest, or be middling about borderline coffee-tea.

Furthermore, Edgington gives us reason to think that the unknowability of a fact does not make a difference (Edgington 1997, p. 314). One can prefer to be given coffee to tea without being able to tell what it is, but that does not imply that, one must prefer borderline coffee-tea to tea.

I do not dispute Edgington's data. At this point in her article she is distinguishing her own position from epistemicism. The point of the above example is to separate being uncertain whether or not a is F from a being a borderline case for an F. For Williamson, there is a true answer to the question of whether a is F. Vagueness shows itself as our (necessary) ignorance of this truth (whether a is F). This differs from the picture I have provided on which uncertainty in genuine borderline cases is metalinguistic. That is not to say that epistemicists can't have a form of metalinguistic uncertainty too. Metalinguistic uncertainty in the presence of a semantic fact of the matter is uncertainty about where the extensions of one's terms are (what the truth-conditions/meanings of one's terms are). That is not the metalinguistic uncertainty that we have on my account. All of the meaning facts can be available, and nonetheless, we can still be uncertain how to use our words.

This notion of metalinguistic uncertainty (uncertainty about use in the absence of a fact about extension) is not one that is available to an epistemicist. With this distinction, we will be able to: (i) retain the force of Edgington's example against epistemicism. (ii) allow an epistemic reading of verities without falling foul of Edgington's example.

(i) Epistemicism is especially targeted by Edgington's example. Epistemicism holds that ALL cases of (apparent) indeterminacy, are really cases of uncertainty. The force of Edgington's example is to indicate that this may not be right. The main idea is that, if indeterminacy is uncertainty, then a situation in which it is uncertain whether ϕ should affect one's (reasons for) actions in the same way as situation in which it is indeterminate whether ϕ. Being uncertain is a reason for going to see colleague C before D and D before E. Indeterminacy need not be. Hence ϕ being uncertain is not the same as ϕ being indeterminate. Epistemicists are certain of always getting one thing or the other (even if they don't or can't know which). Furthermore, not being able to know which does not make a difference. Epistemicists can't explain why uncertainty about what something is makes one act differently to something being a borderline case.

(ii) The metalinguistic uncertainty view does not see all cases of indeterminacy as only apparent. There can genuinely be indeterminate matters. However, indeterminacy arises due to irresolvable metalinguistic uncertainty. Indeterminacy plays a different role in reasons for action than uncertainty about the world, because uncertainty about the world is not the same as metalinguistic uncertainty. As demonstrated by Edgington's example, uncertainty about the world can influence our behaviour in how we interact with the world (whose office should I visit first?). Metalinguistic uncertainty can influence our behaviour in how we use our language (should I use 'tall' to describe her or not?). The switch, from D having a 50-50 chance of having coffee, to D having only borderline coffee-tea, makes a difference to our preferences because the former affects what we do about obtaining coffee, while the latter affects what we say (or fail to say) when handed a cup of unidentifiable liquid.

8.1.3 Verities and Cognitive Efficiency

If indeterminacy arises from irresolvable metalinguistic uncertainty, then borderline cases provide a link between the rules of inference for reasoning in conditions of uncertainty and reasoning in conditions of INDETERMINACY. In addition to the kinds of structural similarities that Edgington highlights in (Edgington 1997),[5] this is an additional reason to take the two to correspond in their logical structure.

The question is where this logic should be applied. Do we want to say that all metalinguistic uncertainty is a form of indeterminacy? For example, we are unlikely ever to be completely certain of a classification. However small, a little bit of doubt might always remain. If we are 0.99 certain that we should classify an object as, say, 'red', and if we cannot be more certain (the uncertainty is irresolvable), should we accept that the matter is slightly indeterminate? Should we even say that the truth value of a statement describing the object as red should get a value of 0.99?

Here, taking *cognitive efficiency* into account suggests that the answer to this question needs to be nuanced. The picture I have been describing of how we interact with a world of messy data leaves us with uncertainties pretty much everywhere. Keeping track of all these uncertainties (even if not consciously), would be a complex and inefficient way to operate. If low levels of uncertainty tend, overwhelmingly, to be good enough (and reliable enough) to support successful communication, then, under pressure of cognitive efficiency, perhaps we simply ignore low levels of uncertainty and operate as if we were dealing with a limit case. Reasoning becomes much simpler when values are at the limit. For many tasks, to simplify cases to the limit will do, and this will not lead us into error. Then again, not all cases will be safe. Edgington rightly draws attention to the lottery paradox (Edgington 1997, §2) as an example of how ignoring small levels of uncertainty can quickly accumulate over repeated inferences.[6]

Perhaps this marks how, in lightening our cognitive load, we become susceptible to losing track of reasons. If we are 0.99 certain that the swatch should be described as red, we have a weighty reason for calling it so. We might also have (slightly less) weighty reasons for classifying a different

[5]These include the facts that both can come in absolutes and degrees, both can be modelled using the real numbers in the range $[0, 1]$, both can be context dependent, both seem to yield only approximate equalities, and neither seem to yield precise mappings (even if they can be modelled with such).

[6]This idea, put in terms of verities, is what lies behind Edgington's treatment of the Sorites paradox. Full details of this will be given in the next chapter.

shaded swatch as red, too. Our reasons for accepting a conjunction will decrease over conjuncts, and our tendency towards cognitive efficiency will put us at risk of misclassification.

There are, then, two ways in which we might view verities. The first will simply be what we reason with in cases of irresolvable metalinguistic uncertainty (indeterminacy). When faced with two borderline cases, we need to reason about them, taking into account their slight differences. One may be slightly less borderline than the other, without giving us sufficient reason to classify either as one way or another. Verities can be seen as the things that we reason over in cases of irresolvable metalinguistic uncertainty. Put another way, these verities only enter into our reasoning when we are faced with irresolvable semantic indecision. This effectively treats verities as the (degrees of) truth values that underpin reasoning in borderline cases. The second way to view verities is simply as a record of (possibly mounting) uncertainty. Given our cognitive tendency to simplify cases to the limit and make Boolean judgements, verities will simply mark or record how much risk we run of unwittingly misjudging a case; of how much danger we are in of taking something to be true when we no longer have sufficient reason to do so.

The two interpretations are not in tension if one accepts the reduction of indeterminacy to irresolvable metalinguistic uncertainty. However, the former is more restricted to application in borderline cases, whereas the latter is a broader and more ubiquitous phenomenon.

For the rest of this chapter, I will defend the first reading of verities. I will argue that, as a logic, verity logic is appropriate for reasoning about borderline cases. In the next chapter the second understanding will be revisited. There it will play an important role in analysing Sorites.

8.2 Defending a Non-Classical Notion of Truth

Interpreted as above, Edgington's Bayesian Logic has a non-classical notion of truth, but it sustains a classical theorems. The task of this chapter is to try to dispel the force of a variety of arguments that are aimed (to varying degrees) at the Bayesian approach. Some of the arguments presented are aimed at fuzzy logic accounts, others at supervaluationism, and still others at Bayesianism in particular. Bayesian approaches share some of the features of both fuzzy logic and supervaluationism. Therefore some arguments specifically targeted at fuzzy or supervaluationist approaches will also be relevant.

I shall present Bayesian responses to all of these arguments. First (§8.2.1) we will consider arguments that specifically target probabilistic logics. Then we look at other arguments that target non-classical and/or non-traditional approaches in general (§§8.2.2-8.2.3).

8.2.1 Against Bayesianism

Fuzzy Criticisms

Bayesianism has been discussed in the literature by those defending fuzzy views (e.g. (Schiffer 2003), (MacFarlane 2009b), (Smith 2008)). One of the major motivating factors for fuzzy theories is supposed to be that they capture intuitions regarding the truth values of conjunctions and disjunctions in cases where there is supposedly no uncertainty, only vagueness. Classical probability logic is supposed to produce wildly unintuitive results in such situations and so, if one wishes to incorporate degrees into one's account, fuzzy logic is the one to choose. MacFarlane has expressed the thought:

> "It seems perfectly appropriate to endorse the conjunctive proposition that Jim is tall and bald and smart to about the same (middling) degree as we endorse the conjuncts separately. Certainly it seems wrong that we should be confident that Jim doesn't have all three properties.
>
> If you don't have these intuitions, try increasing the number of independent properties. With seven independent properties, your credence that Jim has all of them should be less that 0.01, and your credence that Jim doesn't have all of them greater than 0.99." (MacFarlane 2009b, p. 444)[7]

Some writers invoke different notions of belief (Schiffer 2003), and others, different contexts (Smith 2008), to explain the difference between when to employ probabilistic reasoning and when to employ fuzzy logic. When we are in a situation filled with uncertainty, probabilistic reasoning is needed. Two fair coins being tossed have a 0.25 chance of both being heads. Similarly, so the thought goes, when we are in vague situations, we should calculate the degrees of belief/truth we have in terms of fuzzy logic.

[7] MacFarlane cites the point about conjunction as being from (Schiffer 2003, ch 5).

One possible move is to try to undermine the intuitions at play in the vague cases. Famously, many of us are not fully rational when it comes to certain kinds of situations involving chance. Perhaps fuzzy intuitions about conjunctions of independent propositions spring from the same source. Indeed, work by Tversky and Kahneman (Tversky and Kahneman 1974) suggests that intuitions can easily be wrong when it comes to subjective chance judgements.

Smith motivates his fuzzy intuitions with a special kind of vague betting example:[8] Vague bets are different from regular bets in that one wins a proportion of the money based on how true one's bet is: "If one bets on S and S is n true, then one receives n times the stake (Smith 2008, p 242 fn 46). For example, if I bet £1 that Jim is bald and he turns out to be a borderline case of bald (0.5 bald), I should win 50p. If he turned out to be fairly baldish (0.8 bald), I should win 80p. Smiths intuition is that if I bet £1 that Jim is bald and thin, and he is 0.5 bald and 0.5 thin, I should still win 50p.

To see where this reasoning might be going awry, consider a related example of a competition. In the standard non-vague competition, each entrant must decide whether to bet that Jim is bald and thin, bald and not thin, not bald and thin, or not bald and not thin. Only one contestant can choose each of the four combinations, and so, in this competition, there are at most four entrants. There is a grand prize fund of £1 to the one who gets it right. Like vague bets, one gets a proportion of the prize fund based on how true one's bet is. In this case, where Jim is not borderline bald and not borderline thin, one contestant takes home all of the prize money. Suppose that I bet that Jim is bald but not thin. If he is not bald and/or thin, I win nothing. Only the contestant that bets right on both options gets the pound.

Now for the vague competition. Suppose that Jim is 0.5 bald and 0.5 thin, on the fuzzy account, if I bet that he is bald but not thin, I should get 50p, after all, I'm half right. But the same reasoning can't apply to all the contestants without bankrupting the competition organisers. If the prize fund is £1, we can't ALL win 50p. None of us is more right or wrong than the others, so aside from saying that the competition organisers, in order to make a more exciting competition, should have chosen a different individual to bet on, we can say that the fair result is that all four of us go home with 25p each. Bayesian calculus gives exactly this result. If Jim is 0.5 bald and 0.5 thin, (assuming that baldness and thinness are independent of one

[8] He attributes the idea of a vague bet to a presentation given by Peter Milne.

another), he is 0.25 bald and thin.

For the fuzzy theorist, this kind of example has something like the structure of a reverse *Dutch Book* scenario.[9] If you have only £1 to give away and you allow only one bet on each option, you are in danger of paying out more than £1. Under standard probabilistic reasoning, you'll never go bankrupt.

However, perhaps this is an uncharitable interpretation of what vague competitions should be. Perhaps, rather than being a multiplier on the prize fund, one's degree of truth should merely reflect the RELATIVE proportion of the fund, such that how much one gets depends on other people's results. Another argument might be needed.

A similar idea is just to think about ways to be wrong. Fuzzy theorists claim to have very strong intuitions that we should not have a very low confidence that Jim is tall and bald and smart when Jim is borderline tall, borderline bald and borderline smart. However, with n conjuncts, there are $2^n - 1$ other ways to be right or wrong to some degree. Why shouldn't the number of other ways to be wrong be taken into account when determining how true our conjunction is?

Another diagnosis of the problem might be that, in general, conjunctions where both conjuncts are completely independent are something of an oddity. If truth is restricted to Boolean values, the truth conditions of such conjunctions is tolerably clear. But when degrees are involved, it is far less transparent. Intuitions become much stronger when considering dependent cases. These will be addressed in §8.2.2.

Keefe's Criticism

Keefe (2000) has two criticisms of Edgington's verities. However, the first seems to rest on an uncharitable reading, and the second on a failure to fully take into account that the axioms governing credence (modelled with classical probability theory) and the axioms governing verity have a one-to-one correspondence.

Keefe's first criticism centres around Edgington's use of the word 'decide'. When describing conditional verities in (Edgington 1992, p. 202), Edgington says, "Explain $v(A \to B)$ [or $v(B|A)$] as the value to be given to B if we

[9] In a gambling sense, Dutch Books odds are fixed to guarantee money for the bookie when bets are distributed across all options. Here, the reverse is so. In a philosophical sense, one can see the example as a synchronic reverse dutch book. In synchronic dutch books, one can make a series of simultaneous bets and be guaranteed a loss. Here, those entering the competition collectively get a net win.

decide that A is true." In Edgington (1997, p. 306), the wording is slightly different: "We can hypothetically decide to count [a borderline case] as above the borderline—as definitely red—and see what consequences this hypothetical decision has for other propositions." Keefe's first complaint is that "definite truth is not just something we can decide" (Keefe 2000, p. 99). This is uncharitable. In the above passages from Edgington's papers, I take her to simply be giving an informal description of what it is to conditionalise one thing on another, on what the 'given' means in conditional formulas. Perhaps, 'assume' would be more accurate than 'decide'. No serious challenge should be raised on that basis.

Conditional verities are, in some sense, comparable to undischarged assumptions in natural deduction. Suppose that we have as a premise $\beta \to \alpha$, we can perform the following reasoning:

1	$\beta \to \alpha$	Premise
2	$\quad \beta$	Assumption
3	$\quad \alpha$	1, 2 MP

α is true given β. The conditional premise stands for the (logical) relation between the truth of β and the truth of α (it could equally be a different connection such as $\beta \to \neg\alpha$). We conclude that α is true, not *simpliciter*, but true given β. All this can be so whether α is true or false. Conditional verities are similar, however for them, α can be true given β whether α is (definitely) true, (definitely) false or (definitely) indeterminate.

One option that Keefe provides as an interpretation of conditional verities is about right, however. One way to think about the hypothetical decision is to think of hypothetically changing the meaning of the classifier. On the view of meaning I have defended, this amounts to assuming that the usage of the classifier were such as to make some less than clear case into a clear one.[10]

This thought about meaning change is used in Keefe's second objection. It concerns the uniqueness assumption of verity valuations (that formulas, presumably relative to some context, have only one valuation). Suppose

[10]From chapter 6, we could think of such changes, formally, as the shifting of a distribution along the horizontal axis of the graph. For example, a distribution for 'red' could shift further up or down the spectrum of shades. Hypothetically taking some shade to be a clear case would be equivalent to positioning the apex of the curve at that shade.

that two swatches a and b are similar though not identical in colour. Both are borderline cases for a classification as 'red' although b is redder than a. That makes for the following valuation: $v(\texttt{Red}(b) \mid \texttt{Red}(a)) = 1$. Even accepting this, Keefe questions whether a single value could be given for $v(\texttt{Red}(a) \mid \texttt{Red}(b))$. She claims that, since there might be different admissible hypothetical meaning changes for $\texttt{Red}(a)$, there will be no single value for $v(\texttt{Red}(a) \mid \texttt{Red}(b))$. The notion of conditional verity, Keefe claims, does not decide between them and this breaks the uniquieness assumption for the logic (Keefe 2000, pp. 99-100).

However, Keefe's criticism does not take seriously the one-to-one correspondence between verity connectives and the connectives in classical probability theory. For the purpose of her argument, Keefe accepts the initial illustration of a conditional verity (Keefe 2000, p. 99):

$$v(\texttt{Red(b)} \mid \texttt{Red(a)}) = 1$$

Additionally, we have the verity equivalent of Bayes' rule[11] (with the conjunction in the numerator substituted according to the conjunction definition) into which the above value of 1 can be entered:

$$\begin{aligned} v(\texttt{Red(b)} \mid \texttt{Red(a)}) &= \frac{v(\texttt{Red(a)}) \times v(\texttt{Red(b)} \mid \texttt{Red(a)})}{v(\texttt{Red(b)})} \\ &= \frac{v(\texttt{Red(a)})}{v(\texttt{Red(b)})} \end{aligned}$$

This shows that there can only be one value for $v(\texttt{Red(b)} \mid \texttt{Red(a)})$ when we assign non-conditional values to $v(\texttt{Red(a)})$ and $v(\texttt{Red(b)})$. The notion of conditional verities places strict rationality assumptions on the valuations of verities, just as conditional probabilities place rationality constraints on assignments of credences.

[11] Edgington does not cite a verity version of Bayes' rule explicitly, but it follows in the usual manner from two claims: (i) The conjunction rule and (ii) That $v(A \wedge B) = v(B \wedge A)$.

8.2.2 Truth-Functionality

Truth-functionality in logic concerns any complex truth-bearing formula which can be partitioned into at least one truth-bearing formula plus some operator or connective. If the truth-value of a complex is determined entirely by the truth-values of its truth-bearing parts, the connective between these parts or the operator applied to these parts is said to be truth-functional. Even in propositional calculus, a general definition is less than simple, given that both operators such as negation, and connectives such as conjunction and disjunction, are truth-functional. A more general formal characterisation can be given by appealing to the logical theory of types[12]. All well formed formulas can be given a semantic type which is either a basic type or an inductively defined functional type.[13]

Standardly, the two basic types are taken to be e, for entities, and t, for truth-bearers/propositions. Truth-functional functions will either be of type $(t \rightarrow t)$ or $(t \rightarrow (t \rightarrow t))$[14], where their interpretations are fixed across all models. For example, the following defines conjunction:

$$\begin{aligned} [\![\wedge_{(t\rightarrow(t\rightarrow t))}]\!]^M(1) &= \{\langle 1,1\rangle, \langle 0,0\rangle\} \\ [\![\wedge_{(t\rightarrow(t\rightarrow t))}]\!]^M(0) &= \{\langle 1,0\rangle, \langle 0,0\rangle\} \end{aligned}$$

So truth-functionality, in logic at least, concerns logical constants - constants whose interpretations are fixed across all models and are functions of the appropriate type. This definition excludes analogues of connectives such as 'before', since such connectives are not definable as functions taking only truth-values as arguments (the train and the bus can both have left, but the truth of 'the train left before the bus' will still vary in a way that depends on the times that they left). Discussion here will be restricted to the classical connectives of disjunction, conjunction, the classical operator of negation, and to the natural language counterparts for which these connectives provide an interpretation.

Non-classical notions of truth give rise to non-classical notions of truth functionality for the logical constants. For fuzzy-truth there is fuzzy-truth-functionality. For supertruth, there is supertruth-functionality. Variations in the results of non-classical truth functionality have been used as arguments against those positions. They will be the topic of this section, and I will argue that none of them apply to Bayesian logic.

[12] See, for example, (Carpenter 1997).
[13] I gave an example of this in chapter 6.
[14] Or potentially further iterations.

Other arguments derive failures in truth-functionality for non-classical connectives. This is what Williamson calls '*elusiveness*'. Elusiveness will be the topic of the next section (§8.2.3).

Before taking up the details of arguments from truth-functionality, I will go into more detail as to how truth-functionality operates for another non-classical notion of truth, and how a Bayesian notion of truth compares to this. Fuzzy logic approaches are committed to a degree-theoretic account of truth and a degree-functional account of logical constants (Zadeh 1965), (Goguen 1969). Degrees of truth are a measure of degrees of membership of a set. If something is reddish, it is only a member of the set of red things to some degree. Truth comes in degrees, and the degrees of truth of disjunction and conjunction are provided entirely from the degrees of truth of their arguments. Negation is degree functional, but the value of a negation utilises the maximum value of 1. Fuzzy negation, conjunction and disjunction are defined as below (repeated from §1.7.1). Square braces mark fuzzy interpretation:

$$\begin{array}{llll} (\neg_{FL}) & [\neg\alpha] & = & 1-[\alpha] \\ (\wedge_{FL}) & [\alpha \wedge \beta] & = & min\{[\alpha],[\beta]\} \\ (\vee_{FL}) & [\alpha \vee \beta] & = & max\{[\alpha],[\beta]\} \end{array}$$

Fuzzy negation ensures that the sum of the values for a formula and its negation always sum to 1. The degree to which an object is a member of a fuzzy set and the degree to which it is a member of the complement of that set sum to one. Conjunction includes the operation min which returns the value of the lower value of the two conjuncts, hence contradictions ($F(a) \wedge \neg F(a)$) only receive a value of 0 if a is entirely a member of the set for F or entirely a member of its complement set. Disjunction contains the operation max which returns the higher value of the two disjuncts, hence tautologies such as $F(a) \vee \neg F(a)$ only receive a value of 1 if a is entirely a member of the set for F or entirely a member of its complement set.

Conditional values are defined by the verity equivalent of Bayes' Theorem:

$$v(\alpha \mid \beta) \quad = \quad \frac{v(\alpha \wedge \beta)}{v(\beta)}$$

Given that, for dependent α and β, the definition of $v(\alpha \wedge \beta)$ contains either $v(\beta|\alpha)$ or $v(\alpha|\beta)$, conditional values are to some extent basic. They are not definable purely in terms of $v(\alpha)$ and $v(\beta)$, so definitions of $\wedge_{BL}$ and $\vee_{BL}$ are not purely degree functional. However, in the set of special cases wherein the degree of truth of α has no effect upon the degree of truth of β and *vice versa*, conditional values are reducible to the values of their atomic formulas such that $v(\alpha \mid \beta) = v(\alpha)$ and $v(\beta \mid \alpha) = v(\beta)$. We thereby get a derived set of formulas for this special case of Independent Bayesian Logic (IBL) in which the functions for connectives are complex, but nonetheless degree functional:[15]

$$\begin{array}{lllll} (\neg_{IBL}) & v(\neg\alpha) & = & 1 - v(\alpha) \\ (\wedge_{IBL}) & v(\alpha \wedge \beta) & = & v(\alpha) \times v(\beta) \\ (\vee_{IBL}) & v(\alpha \vee \beta) & = & (v(\alpha) + v(\beta)) - (v(\alpha \times \beta)) \end{array}$$

Values for conjunction and disjunction are calculated differently from fuzzy systems, but both are Boolean at the limit.

In many other cases, the (independent) Bayesian and fuzzy accounts will not produce either Boolean or commensurate results. If A has a value of 0.8 and C has a value 0.4, the fuzzy conjunction and disjunction of A and C will be 0.4 and 0.8 respectively, whereas the IBL values would be 0.32 and 0.88. However, insofar as both accounts are degree functional, one might suppose that any arguments against one could be an argument against the other. Yet typically, and for good reason, arguments against fuzzy connectives tend to include examples where the values of statements are not independent.

For example, from the definitions of $\wedge_{FL}$ and $\vee_{FL}$ unintuitive results are obtained when the values of the conjuncts or disjuncts are not at the limit (see for example (Edgington 1997), (Fine 1975, pp 269-70), (Williamson 1994, pp 135-138)). The law of the excluded middle (LEM) and the law of non-contradiction (LNC) both fail when $0 < [\alpha] < 1$ since $[\alpha \vee \neg\alpha] < 1$ and $[\alpha \wedge \neg\alpha] > 0$. Williamson gives a linguistic example for conjunction and disjunction. In vague cases, the same or similar values might be given for ‘He is awake’ and ‘He is asleep’. But since no-one can be both asleep and awake, their conjunction should yield 0, not whatever the lower of the two values are. Equally, since being awake and asleep are complementary states, that someone is awake or asleep should be completely true. But under fuzzy logic, the value for the disjunction is just the greater of the values of the disjuncts.

If the functions for connectives were only meant to cover independent cases, it could be argued that a disservice is being done to fuzzy systems here. To counter this response, we need a clearly unintuitive case of an

[15]Probabilistic equivalents are shown to be consequences of the Kolmogorov axioms in (Kolmogorov 1950).

independent conjunction and an independent disjunction. For the case against fuzzy disjunction, suppose α has a value of 0.9, and we wish to evaluate the value of different formulas disjoined with α. If β is valued at 0.9, $\alpha \vee \beta$ will be 0.9 which might seem *prima facie* reasonable. However, if γ has a value of only 0.1, $\alpha \vee \gamma$ will still be 0.9 and one might think that there should be a difference between the two cases. We can grant that, if two formulas are independent, adding a disjunct to one shouldn't decrease its value. For the same reasons, the classical inference rule of disjunction introduction is valid. However, we might still allow that adding a disjunct could INCREASE the value of the formula. The difference between the two above cases can therefore be seen as bringing into doubt the *prima facie* reasonableness of the value of $\alpha \vee \beta$ being 0.9 when the values of both α and β are 0.9. Put another way, there is an intuition that the truth of a disjunction should be at least as true as its truest disjunct, not just the same degree of truth as its truest disjunct. If that intuition is granted, then we would expect the value of $\alpha \vee \beta$ to be $0.9 + n$ and the value of $\alpha \vee \gamma$ to be $0.9 + m$ such that $n > m$. Fuzzy logic goes against this intuition. IBL gives one exactly that.

A similar, argument can be presented against fuzzy conjunction. Fuzzy conjunction rightly assumes that conjunction can never result in an increased degree of truth. But one might think that, depending on the conjunct, the value may be lowered to different degrees. For example, if α has a value of 0.1, β a value of 0.9, and γ a value of 0.1, one expects $[\alpha \wedge \gamma]$ to be lower than $[\alpha \wedge \beta]$. Fuzzy logic denies this, Bayesian logic explains it.

Degree functionality seems hopeless, but not all degree theories are degree functional. IBL is degree functional, BL is not fully degree functional. However, Williamson signposts a warning for non-degree functional degree theories:

> "To deny degree-functionality is not to deny that truth is a matter of degree. Conjunctions, disjunctions and conditionals might be true to degrees not determined by the degrees of truth of their components. But the failure of degree-functionality does make it hard to see how degrees could be central to semantic theory." (Williamson 1994, p. 138)

However, BL seems to be a case where degrees are essential semantically (in giving values of complexes). For example, conditional degrees reflect logical connections between formulae. In BL, functions for connectives are defined from two types of inputs: (i) unconditional degrees of statements, and (ii) conditional values (which model the logical structure between statements). The values of statements and a representation of their logical connections are both central to semantic theory.

Arguments against fuzzy logic have no impact on BL. The functions for IBL connectives come straight out of the functions that define BL connectives. Employing the latter in the awake/asleep example gives exactly the right results. The general formula for calculating a conjunction of $A \wedge B$ is the product of the value of A with the value of B, given A.[16] For independent cases, the value of B, given A is equivalent to the value of B. For dependent cases such as 'He is awake and he is asleep', the value of 'He is asleep, given he is awake' will be 0 thus ensuring that the value of the whole is 0.

From all this we have a couple of reasons to favour Bayesian accounts over fuzzy ones. In independent cases, fuzzy logic doesn't seem to be able to account for the difference between conjoining or disjoining formulas with greater or lesser degrees of truth. Furthermore, whereas Bayesian accounts move from formulas for dependent cases to functions for independent ones easily, it is not clear how a fuzzy account could be extended to cover dependent cases. Insofar as Bayesian logical constants are degree functional, they avoid the problems faced by fuzzy logical constants. The question remains whether the less than total degree functionality of Bayesianism (arising from conditional values) gives rise to alternative problems. Supervaluational logic is not supertruth-functional. Williamson (1994) uses this as a foundation for his remarks about the elusiveness of supertruth. These remarks will be evaluated next.

8.2.3 Elusiveness

The term 'elusiveness' is first used by Williamson in relation to the lack of supertruth-functionality of supertruth:

> "According to supervaluationism, 'p or q' is sometimes true when no answer to the question 'Which?' is true. For similar reasons, 'Something is F' is sometimes true when no answer to the question 'Which thing is F?' is true. In this sense, supertruth is elusive." (Williamson 1994, p. 153)

Complaints of elusiveness are, in effect, just complaints about a failing of truth-functionality. We could, therefore, see it as a clear choice that a

[16]Or it's equivalent, the value of B times the value of A, given B.

theorist has in their theory of truth and logic – to adopt or not to adopt truth-functionality. The choice might not always be quite so clear however. We saw in the last section that, Bayesian logic does not fall cleanly on either side of this divide. First, Williamson's complaints about elusiveness will be detailed. Then, I will discuss the extent to which these concerns can be applied to a Bayesian logic.

Supervaluationism does not have a degree based notion of truth in the way that fuzzy logic does. As a result, it applies equally well to independent and dependent cases, and both LEM and LNC are preserved.[17] Supervaluations work on the notion of there being admissible precisifications for the content of any utterance that contains a vague term. Full details can be found in chapter 1 (§1.6). Here is a quick summary. Admissible precisifications are determined by the facts of the language. Clearly non-red precisifications are not admissible precisifications for the term 'red'. Suppose that a predicate F has three admissible valuations, F_1, F_2, and F_3. For a term referring to an object in the domain a, Fa is supertrue when F_1a, F_2a, and F_3a are all true, superfalse when F_1a, F_2a, and F_3a are all false, and neither true nor false on other combinations.[18] Precisifications are evaluated classically. They adhere to bivalence, excluded middle and non-contradiction. LEM is preserved for supertruth/falsity because at any sharpening F_n, and for an arbitrary object a, it is true that $F_na \vee \neg F_na$. Hence $Fa \vee \neg Fa$ is supertrue. LNC is preserved, since at any sharpening F_n, and for an arbitrary object a, it is true that $\neg(F_na \wedge \neg F_na)$. Hence $\neg(Fa \wedge \neg Fa)$ is supertrue.

The values of disjunctions are not supertruth-functional however. $Fa \vee Fb$ can be supertrue when neither Fa nor Fb is, provided that for any sharpening i, either F_ia or F_ib is true. Equally, $\exists x.Fx$ can be supertrue when no instance of Fa is supertrue provided that for every sharpening of F, at least one entity term truly applies to that sharpening.

As Williamson points out, elusiveness troubles also stretch to quantification (as in the short Sorites) (Williamson 1994, §5.4). Boundary drawing existential statements are supertrue because under every precisification, 'n grains make a heap, but $n+1$ grains do not' is true for some n. However, no instance of 'n grains make a heap, but $n+1$ grains do not' is supertrue. I will now consider whether elusiveness is a problem for BL.

[17] Even though bivalence is not (more on this below).

[18] Precisifications are, however, connected such that $F_1a \models F_2a$ and $F_2a \models F_3a$. Therefore, the combination of F_1a and F_3a being true when F_2a is false is out, as is the combination when those truth-values are reversed.

LNC is preserved in BL. The conditional values $v(\phi \mid \neg\phi)$ and $v(\neg\phi \mid \phi)$ will always equal 0. If $v(\phi) = n$, a conjunction of ϕ and $\neg\phi$ will always be $n \times 0 = 0$. Because $v(\neg\phi) = 1 - v(\phi)$, the negation of the conjunction will be 1. LEM is preserved in BL too. A disjunction of ϕ and $\neg\phi$ will always be $n + (1 - n) - 0 = 1$. Even though any instance of LEM or LNC will have a value of 1, neither ϕ nor $\neg\phi$ needs to have a value of 1. In that sense, arguably, BL is elusive.

However, there is some reason to think that BL elusiveness is less troubling than supervaluational elusiveness. As already noted, BL values are given by functions on semantically central (conditional and unconditional) values (§8.2.2). Furthermore, conditional values are no more or less suspect than undischarged assumptions in natural deduction (§8.2.1).

A conjunction can only get a value of 1 if (i) The value of at least one conjunct, α, is 1, and (ii) The value $v(\beta|\alpha)$ is 1. There is no elusiveness there. A conjunction can only get a value 0 if either (i) The value of at least one conjunct, α is 0, or (ii) The value $v(\beta|\alpha)$ is 0. This could be where elusiveness threatens, since neither $v(\alpha)$ nor $v(\beta)$ needs to be 0 for the conjunction to be 0. However, degree functionality of BL connectives is a functionality of two different kinds of degrees: unconditional degrees, and conditional degrees. The falsity of a conjunction can result, either because one of the conjuncts is false (this is reflected by unconditional degrees), or because both conjuncts cannot be true together (this is reflected by conditional degrees). Verity conjunctions have a shade of complexity that Boolean conjunctions do not. Verity conjunctions are complex, but they are not elusive.

The 'problem cases' of disjunction and existential quantification raised by Williamson must still be addressed. I take them in turn. There are two ways for a disjunction to receive a value of 1: First, when at least one disjunct has value of 1, but the two values are independent (so that for all values of β, $(1 + v(\beta) - (1 \times v(\beta)) = 1)$). Second, if conditional values make a difference, the disjuncts can be any values provided that their sum is greater than or equal to 1 and that they stand in the right relationship conditionally. In the first case, truth is not elusive as by definition, one disjunct is completely true. Other cases are less straight forward, and instances of LEM are where the pressure of elusiveness is felt most acutely. Because LEM holds in the system, $v(\alpha \vee \neg\alpha) = 1$ for any value of $v(\alpha)$. Apart from the case where the value for α or its negation is 1, the disjunction has a value of 1 even though neither disjunct does. In this sense, truth in BL could be elusive. However, we can again put things in terms of the two kinds of degrees that are of semantic importance. LEM disjunctions can get a value of 1 because the negation in one disjunct will always ensure their unconditional values sum to 1. But also, the formula for disjunction contains a conjunction (including a conditional verity value). For statements of LEM, this conditional value is 0 which indicates that the disjuncts cannot be true together. Saying that they cannot be true together amounts to saying that the extent to which one is true is the extent to which the other is false. Truth is shared over the disjuncts. Disjunctions are, again, complex, but they are not elusive.

A simple story for existentials can be given by merely evaluating them as the disjunctions of their instances.[19] On this assumption, what applies to a disjunction applies to them.

8.3 Bivalence and Disquotation

The principle of bivalence, that every utterance is either true or false, is violated on both fuzzy and supervaluational accounts. In fuzzy accounts, even LEM doesn't hold. On supervaluational accounts, LEM holds. Every utterance is interpreted as true or false on every valuation, and so for any valuation i, $\alpha \vee \neg\alpha$. However, since supertruth requires truth on all valuations and not every sentence is supertrue or super-false, supertruth is not bivalent.

Independently of the role of T-schemas in Tarski (see chapter 2), truth has been argued to be disquotational. For example, Williamson holds that if there is any minimal thing one ought to be able to say about truth irrespective of theory, it is that it is disquotational (Williamson 1994, §5.7). In other words, for any natural language utterance u, if u says that P, we can disquote: u is true iff P. There is definitely something plausible about the disquotational property of truth. Indeed, it is meant to be intuitively justified by an understanding of the truth predicate.

Williamson argued that from principles of disquotation, LEM, and a denial of bivalence one gets a contradiction (full details in §1.4.2).[20] The force of Williamson's argument stems from the fact that LEM and disquotation principles imply bivalence (where the left and right sides of disquota-

[19]A full story for quantification in BL needs to be given. A promising proposal has very recently been made in (Cooper et al. 2013).

[20]That is why Williamson's argument does not affect intuitionism. Intuitionist logic does not support LEM.

tional principles are taken to be semantically equivalent):

($D1^+$)	u is true iff P.
($D1^-$)	u is false iff Not:P.
(LEM1)	P or Not:P.
	Substitute 'u is true' for P.
	Substitute 'u is false' for Not:P.
(B1)	u is true or u is false.

The approach that I have defended supports LEM. Clearly, I must say something about disquotation and bivalence.

Disquotation

The standard disquotational truth schema is ($D1^+$). One question we can ask is how ($D1^+$) can be understood in terms of degrees. For 'is true', we could mean 'u is true to some (i.e. any) degree', 'u is completely true', or 'u is true to degree n':

(D2*)	u is true to some (i.e. any) degree, if and only if P
(D3*)	u is completely true if and only if P
(D4*)	u is true to degree n if and only if P

Ignoring P for a moment, notice that, when put in degrees, we do not need a degree version of ($D1^-$). In our degrees talk, we have not, so far, discussed falsity, but degrees of falsity are directly translatable into degrees of truth:

$$\text{If } u \text{ is true to degree } n \text{ then } u \text{ is false to degree } 1 - n$$

The tension between degrees and disquotation arises when we ask what P is. Take Williamson's above prefix of disquotation "If u says that P". When placed before (D2*) or (D4*), the principle sounds wrong.[21] On the view defended, I have argued for a greater separation of meaning and truth conditions. If the "says that P" is, in some sense, supposed to capture the meaning of u, then we ought to reject it being placed as a statement of u's truth condition.

Instead of P being what u says, we can look at an alternative in the

[21] It is less odd for (D3*): If u says that P, u is completely true if and only if P.

literature. Disquotation has also been understood is in terms of assertion. For example, Quine puts the point in terms of calling and affirmation:

> "By calling the sentence true, we call snow white [...] We may affirm the single sentence just by uttering it unaided by quotation or by the truth predicate." (Quine 1970/1986, p. 12)

To find a suitable substitution for P, we can provide an equivalence for what it is to assert that u is true. I have argued that truth is connected to there being reasons for belief. Reasons are not subjective, they are in part provided by how words are used and in part provided by other information contained in the world (modelled with resource situations). I propose that an assertion of 'u is true' is equivalent to STAKING A CLAIM for there being reasons to support an assertion of u.

Just as it can be reasonable to believe something to a greater or lesser degree, there can reasons to support an assertion of u to greater or lesser degrees. This follows from the account of meaning that has been proposed. For example, if the information carried by 'red' is that it correlates strongly with shade a and weakly with shade b, then if an object is a, we have a greater reason to call it red, than if it were shade b.

Call the degree to which there are reasons to support an assertion of u, u-*entitlement*. That means that u-entitlement is just another way of referring to the amount of metalinguistic certainty there is for an utterance of u.[22]

The notion of u-entitlement is what, I suggest, provides a plausible logically equivalent substitution for P in our disquotation schemas:

(D2$^+$) u is true to some (i.e. any) degree, if and only if there is some u-entitlement.

(D3$^+$) u is completely true if and only if there is u-entitlement of 1.

(D4$^+$) u is true to degree n if and only if there is u-entitlement of n.

In effect, these say that u is true if and only if one is entitled to assert u.[23] This is, admittedly, not pure disquotation, but it does, I contend, provide a

[22]Recall that this is independent of what reasons an agent knows about or is aware of. Just as truth is not a relative notion, entitlement to assert or utter, on this understanding, is not relative either. Given some described situation, some discourse situation, and possibly some resource situation(s), there will be only one degree of entitlement for an assertion/utterance of u. However, u-entitlement can vary with a variation in those situations.

[23]It is important to keep in mind both that entitlement is metalinguistic (it concerns how words are used), but also that all the information required to be able to know if one is semantically entitled to assert needn't be accessible to the agent. As pointed out by an examiner, on some understanding of entitlement to assert, one may have no entitlement to assert 'the coin will land heads', and yet that assertion may be true. I do not disagree,

plausible link between truth and reasons for linguistic actions.

The above principles weaken the Quinean identity (to assert u is true IS to assert u), to the claim that, to assert u is true is to stake an entitlement to assert u. Doing this has other benefits. For example, we can explain why sometimes asserting that u is true is inappropriate when asserting u is not. It would be very odd to say: "'I'm going to the park tomorrow' is true. Would you like to come?". Plausibly this is because, in such a construction, there is little need to state an entitlement to assert one's own intentions.[24]

As a final note on disquotation, we should consider one of the philosophical tasks that has be loaded upon such principles ((Quine 1970/1986), (Field 1994), (Horwich 1990)). This will test our new versions to some extent. One role for disquotation is to allow us to express infinite conjunctions/disjunctions etc. For example, if Alex wants to express that she believes anything that Charlie says, then rather than state this as an infinite conjunction (If Charlie says "Grass is green", then Grass is green, and if Charlie says "Snow is white", Snow is white...), we can use the disquotational schema to reinterpret the conjunction (If Charlie says "Grass is green", then 'Grass is green' is true, and if Charlie says "Snow is white", then 'Snow is white' is true...). Then we can quantify over the sentences: 'For all p, if Charlie says "p", then 'p' is true, which is supposed to capture that Alex thinks that everything Charlie says is true.

A similar story can be given using the above principles, except that the first infinite conjunction must be rephrased. One way to describe what Alex's state is like when Charlie says "Grass is green" is that she is metalinguistically certain asserting that Grass is green. If that is allowable, then the same infinite conjunction story can be told. We can go from 'If Charlie says "Grass is green", then one is completely entitled to assert that Grass is green, and if Charlie says "Snow is white", then one is completely entitled to assert that Snow is white...', to 'If Charlie says "Grass is green", then 'Grass is green' is true, and if Charlie says "Snow is white", then 'Snow is white' is true...'. Which again allows us to quantify in the same way as

but on the understanding of 'entitlement' in question, this is not a problem case. One does not know which way the coin will land, and so, *a forteriori*, one cannot know what the best way of describing the way the coin lands is. However, say that it lands heads. Then, given the way 'land heads' and 'land tails' are used (given the information they carry), one would not, irrespective of what one knows at the time of utterance, be entitled to assert 'the coin will land tails', but one would be entitled to assert 'the coin will land heads'. In a nutshell, semantic entitlement concerns whether the information carried by words sufficiently matches up with how things are or turn out to be whether or not the agent knows how things are or how they'll turn out out be.

[24]There may, of course, be contexts in which this is possible.

above.

Bivalence

The above disquotational principles can be used, along with a specific statement of LEM, to derive some options for degree theoretic versions of Bivalence. The relevant instances of LEM are about u-entitlement:

(LEM2) There is u entitlement, or Not: there is u entitlement.
(LEM3) u entitlement is 1, or Not: u entitlement is 1.
(LEM4) u entitlement is n, or Not: u entitlement is n

This results in three corresponding principles of bivalence:

(B2) Either u is true to some (any) degree, or Not: u true to some (any) degree.
(B3) Either u is completely true, or Not: u is completely true.
(B4) Either u is true to degree n, or Not: u is true to degree n.

(B2) is acceptable because its first disjunct always is (its second disjunct never is). (B3) and (B4) are always true. What (B3) and (B4) reflect is that when u has some degree of truth, there are many other ways for it not to be true to that degree.

8.4 Summary

I have attempted to achieve two goals in this chapter. The first, was to give an account of indeterminacy as a form of irresolvable metalinguistic uncertainty. Critically, this was not uncertainty about facts (what exact meaning I am currently expressing/how the world really is). It was metalinguistic uncertainty about whether to make a classification using a predicate. Cases where one cannot and could not resolve metalinguistic uncertainty are, I claimed, just those cases where indeterminacy arises from borderline cases. This account of indeterminacy allowed for two interpretations of verities. (i) As degrees of truth. (ii) As records of increasing metalinguistic uncertainty.

Having connected my account of meaning and information with an interpretation of degrees of truth and with Edgington's logic to reason with them, a number of arguments against non-classical positions were discussed. Bayesian logic fares well. The account of truth I have argued for supports versions of the disquotational feature of truth (albeit interpreted as relating to assertion). In turn, the disquotational schemas support versions of principles of bivalence. Furthermore, a major advantage of the Bayesian system over its fuzzy and supervaluational counterparts is that it retains classical theorems.

In the next and final chapter, now we have both a story about meaning, context shift, truth, and logic, we can finally acknowledge the elephant in the room and look at the sorites.

9

THE SORITES

It is finally time to focus on the central topic of the philosophical conception of vagueness: Sorites paradoxes. In chapter 1, I argued that the Sorites paradox should be a test for an account of vagueness, not a motivation for it. I have given a motivation for a probabilistic conception of meaning based on pressures of communication and semantic learning. Now, the Sorites can be deployed as just one way to look for cracks in the account.

The Sorites paradox concerns logic and truth, as well as meaning. Central questions are:

(Q1) Are Soritical arguments valid? If not,

   (a) which inference(s) is/are invalid?

   (b) why do similar inferences appear to be valid/harmless in other applications?

(Q2) If they are valid, are they sound? If not,

   (a) which premise(s) is/are false?

   (b) why is it so hard to spot the false one(s)?

(Q3) How can we know the meanings of words and not know where to dig our heels in on the slippery slope?

These questions tally, roughly, with those put forward by Graff Fara in relation to the Sorites that were summarised in chapter 1 (§1.5.2). Unlike the full version found in (Graff Fara 2000), I do not pose these questions in terms of first-order formulas.

Over the course of this short chapter, I will suggest two compatible solutions to the sorites. One is informal and stated in terms of reasons. The other is formal, and given in Bayesian logic. After discussing these solutions, I will argue that all of the above questions receive a satisfactory answer.

9.1 Solutions

The Sorites can be taken as a puzzle about logic, about semantics, about truth value judgements, or about meanings of predicates. The picture I have given naturally accommodates all of these ways of seeing the paradox. The solutions I will provide, for the most part, will not be particularly original. Given my adoption of Edgington's logic, it should be no surprise that I will adopt her ready made formal solution. The other solution will be one from Charles Travis. His attitude towards vagueness is inspired, in part, by Austin. His main contention accords with the discussion in §7.6. It is not the case that predicates (words) are vague. People are vague in their use of words.

9.1.1 Chains of Reasons

In talking about truth and in putting forward the view that truth (being true) is a pragmatic property of utterances, reasons came very much to the fore. Recall that this view is not relativistic. In science, establishing truth is about weighing evidence and testing hypotheses. As individuals, our semantic competence and our arrival at truth value judgements can be seen similarly. Words carry information that gives us expectations for how the world is. In many situations, it is entirely clear whether or not the world is the way that someone's words have led us to expect. Disagreements on such cases are not relativistic disagreements over truth. One party can simply be wrong. They can be wrong because they are blind to the evidence. Namely, how the words in question are used.

Nonetheless, there can be cases where all the facts are in, and yet we have insufficient reason to form a judgement. Unlike cases of empirical discovery where we tend to assume (rightly) that some hypothesis is unconfirmed for lack of evidence, in special metalinguistic cases, there is no more evidence to find. Everything can be entirely open to view, and yet we can fail to have reason to judge something true or false.

At these special points, two distinct notions come together. Our reasons for judging something true are distinct from whether or not something is true because we can fail to have access to the facts that make something true. Metalinguistic borderline cases connect these two notions because they are cases where we do have access to all the facts and it is no failing in us that these facts do not add up to a truth value.

The pragmatic view of truth fell out of (i) a rejection of truth and truth conditions as necessary parts of meaning, and (ii) the development of an account of meaning that rested on an account of natural, statistical information. (ii) led to the view that a description of something as red carries information. Learning a language amounts to learning to decode and manipulate this information. Understanding the description amounts to forming certain (weighted) expectations for how the world is. (i) and (ii) entailed that we do not have sharp extensions for vague terms.

Knowing the meaning of terms in a language and being competent in the use of language amounts to being able to use the information conveyed by the terms in the language to form judgements for how objects in one's environment should be classified. The meanings of terms do not determine answers to classification problems, but they do provide REASONS for and against classifying things in one way or the other.

Sorites arguments are not merely about simple one-off classifications, or about finding, say, the last red (or first orange) swatch in a fabric sample booklet. The arguments are designed, in long sorites at least, to lead us down a slippery slope. Ignoring the (formal) logical aspects of the argument, and so, sticking just with natural language reasoning, each move down the slope can be viewed as a tallying of a chain of reasons. Here is an informal story about how tracking these reasons might go:

> A looks at the first swatch and forms a judgement: "The first swatch is red". Her reasons for this may be solidly based on her understanding of what 'red' means. In her current circumstances, she has as strong (or stronger) reasons for judging the shade of the first swatch to be red than she would were it some other shade. The second swatch looks very similar (or even identical) to the first swatch. That it looks similar is one reason for A to form the judgement that "If the first swatch was red, then so is the second", or just "The second swatch is red". At this point, A's

understanding of 'red' also backs up this judgement.

Our agent in the story can continue her way down the slope using similar reasons and reasoning. At each point, the similarity of each successive swatch to its predecessor provides her with a reason to judge it as red. However, there are other reasons that A should, but may not be, taking into account. Here is how Travis sees the situation:

> "Again we suppose that the first member is unmistakably red. That, plus indistinguishability, gives us excellent reason to say the same of the next member. Such is transmitted, not much diminished to the next member, and so on. As we go through the series, however, the force of the reason transmitted by this process diminishes. The reason is that a different, though parallel chain of reasons is transmitted from the last member of the series, which is indisputably orange, in the opposite direction. So for any member of the series, there will be some reason, no matter how slight, for calling it red, and similarly for calling it orange. Since the series is long, there may be many cases where one reason clearly outweighs the other. The first and last member are two such. But there may also be many cases where weighing yields no clear result." (Travis 2008e, pp. 221-2)

Travis' account fits well with all that has been said about truth, meaning, and reasons. I would add, simply, that another source of reasons is present too: the meaning of the classifier term itself. Despite looking similar to its predecessor, subsequent swatches will nonetheless start to look less red. They will start to become less and less as one would expect of something described as red. In contrast to the earlier cases, the information carried by the classifier we are employing starts slowly but surely to provide less and less reason to make the next classification.

Gradually, as reasons begin to conflict, we may ask: "Is there a definitive answer to when one reason or set of reasons (sufficiently) outweighs another?" or "Is there a definitive answer to which swatch is the last red one?". The answer to both of these questions, I suspect, is "Probably not". The weighing up of reasons (and whether this is at all possible) is a well studied topic in the philosophy of action. I leave a decision on this matter to future research.

9.1.2 Decreasing Verities

In the last chapter (8.1.3), I discussed two possible interpretations of verities. One, discussed at length in the previous chapter, was restricted to borderline cases and was suggested as a way to interpret degrees of truth. On the other, verities are records of the mounting risk of misclassification. On this second understanding verities mark how much entitlement we really have in our classifications, despite our (usually) unproblematic ignoring of them. We frequently make Boolean truth judgements on matters for which we lack complete and total entitlement. This understanding of verities marks our level of risk. It is this understanding of verities that will be appealed to in modelling the Sorites.

I will give two analyses of the Sorites. First, where tolerance conditionals are interpreted as conditional verities. Second, where tolerance conditionals are treated as material conditionals.

Tolerance Conditionals as Conditional Verities

Very few cases, even in context, are likely to get a verity of 1 on this understanding of them.[1] Each swatch in the series, then, gets a verity value. Say there are 99 swatches ($\mathtt{s}_{99}$ - $\mathtt{s}_1$) such that their individual verities go in 0.01 increments downwards from 0.99 to 0.01. So $v(\mathtt{Red}(\mathtt{s_{99}})) = 0.99$ and $v(\mathtt{Red}(\mathtt{s_1})) = 0.01$. The conditional verity of an earlier swatch being red given that a later one is, can be 1 (or, if not 1, very very high):

$$v(\mathtt{Red}(\mathtt{s_n}) \mid \mathtt{Red}(\mathtt{s_m})) = 1 \quad \text{when } n \geqslant m$$

Tolerance conditionals for this series will be of the form 'If $\mathtt{Red}(\mathtt{s_n})$, then $\mathtt{Red}(\mathtt{s_m})$'. If we represent these as conditional verities that will be:

$$v(\mathtt{Red}(\mathtt{s_m}) \mid \mathtt{Red}(\mathtt{s_n}))$$

which can be calculated using the verity equivalent of Bayes' rule. The result is then simplified by substituting 1 for $v(\mathtt{Red}(\mathtt{s_n}) \mid \mathtt{Red}(\mathtt{s_m}))$:

[1] Although, if there are any cases of indisputable cases of classifications, these might be good candidates. It should also be stressed that this understanding of verities is not Edgington's. On Edgington's view, most statements will receive a verity of 1 or 0.

$$
\begin{aligned}
v(\texttt{Red}(\texttt{s}_\texttt{m}) \mid \texttt{Red}(\texttt{s}_\texttt{n})) &= \frac{v(\texttt{Red}(\texttt{s}_\texttt{m})) \times v(\texttt{Red}(\texttt{s}_\texttt{n}) \mid \texttt{Red}(\texttt{s}_\texttt{m}))}{v(\texttt{Red}(\texttt{s}_\texttt{n}))} \\
&= \frac{v(\texttt{Red}(\texttt{s}_\texttt{m}))}{v(\texttt{Red}(\texttt{s}_\texttt{n}))}
\end{aligned}
$$

On this modelling, the verities of tolerance conditionals will decrease increasingly steeply towards the end of the series (most have a value well above 0.90).[2] The last tolerance conditional has the lowest value: $v(\texttt{Red}(\texttt{s}_1) \mid \texttt{Red}(\texttt{s}_2) = \frac{0.01}{0.02} = 0.5$. If (tolerance) conditionals are modelled using conditional verities (this is the proposal in (Edgington 1992)), we have a very good reason to accept each tolerance conditional until the last few. Even then the minimum is 0.5.

The Sorites could be approximated by a conjunction of all the tolerance conditionals. BL conjunctions of conditional verities are complex, but even on an IBL (product) conjunction, unverities will mount. In fact, when calculated, the decrease is a little less than 0.01 per step. The conclusion has a verity of 0.0101.

Edgington's definition of validity is that the unverity of the conclusion cannot exceed the sum of the unverities of the premises. The Sorites is therefore valid but unsound. Its validity can be demonstrated by showing that, for each step of *modus ponens* in the series, the sum of the unverity of the tolerance conditional and the unverity of the antecedent of the tolerance conditional is greater than the unverity of the consequent. For example, the first step is valid:[3]

$$
\begin{aligned}
&v(Red(s_{99})) && = 0.99 \\
&v(Red(s_{98}) \mid Red(s_{99})) && = 0.9899 \\
&v(Red(s_{98})) && = 0.98
\end{aligned}
$$

The sum of the unverities of the conditional and its antecedent is $0.01 + 0.0101 = 0.0201$. The unverity of the conclusion is 0.02. That means that the inference is valid.

Though valid, the argument is unsound. However, unlike classical unsoundness in which at least one premise is false, BL allows that a

[2] A sample of these are: $\frac{0.98}{0.99} = 0.99$, $\frac{0.78}{0.79} = 0.99$, $\frac{0.58}{0.59} = 0.98$, $\frac{0.38}{0.39} = 0.97$, $\frac{0.18}{0.19} = 0.95$, $\frac{0.08}{0.09} = 0.89$.

[3] All steps will be valid if the decrease in value is small enough.

conjunction can have a 0 or near 0 value even when no conjunct has a low value. This is the case with our tolerance conditionals above. The lowest value there was 0.5, but the IBL conjunction of all is very low.

Tolerance Conditionals as Material Conditionals

In Edgington (1997), tolerance conditionals are treated as material conditionals. The full formal details of this treatment can be found in (Edgington 1997). I will give only a brief sketch here.

On a material analysis, we can understand the negation of any tolerance conditional as the conjunction of the antecedent with a negation of the consequent:

$$\begin{aligned} v(\neg(\texttt{Red}(\texttt{s}_\texttt{n}) \rightarrow \texttt{Red}(\texttt{s}_\texttt{m}))) &= v(\texttt{Red}(\texttt{s}_\texttt{n}) \wedge \neg\texttt{Red}(\texttt{s}_\texttt{m})) \\ &= v(\texttt{Red}(\texttt{s}_\texttt{n})) \times v(\neg\texttt{Red}(\texttt{s}_\texttt{m}) \mid \texttt{Red}(\texttt{s}_\texttt{n})) \end{aligned}$$

The value for the conditional verity can be calculated:

$$v(\neg\texttt{Red}(\texttt{s}_\texttt{m}) \mid \texttt{Red}(\texttt{s}_\texttt{n})) = 1 - v(\texttt{Red}(\texttt{s}_\texttt{m}) \mid \texttt{Red}(\texttt{s}_\texttt{n}))$$

Using the values from the previous case, the negation of every tolerance conditional has a value of 0.01. For example:

$$v(\texttt{Red}(\texttt{s}_\texttt{n})) = 0.99$$

$$v(\texttt{Red}(\texttt{s}_\texttt{m})) = 0.98$$

$$\begin{aligned} v(\neg\texttt{Red}(\texttt{s}_\texttt{m}) \mid \texttt{Red}(\texttt{s}_\texttt{n})) &= 1 - \frac{0.98}{0.99} \\ &= 0.101 \end{aligned}$$

$$v(\neg(\texttt{Red}(\texttt{s}_\texttt{n}) \rightarrow \texttt{Red}(\texttt{s}_\texttt{m}))) = 0.101 \times 0.99$$
$$= 0.01$$

So, because the negation of every tolerance conditional is 0.01, every tolerance conditional has a value of 0.99. This small unverity in each conditional is what makes the unverity of each consequent greater than that of the antecedent. With each step of *modus ponens*, one moves 0.01 closer to 0.

Sorities arguments are valid, but unsound on this modelling too. Although each step in the long sorites is valid (for material conditionals the unverity of the antecedent plus the unverity of the conditional equals the unverity of the consequent). It is, however unsound, not because some one tolerance conditional is false, but because lots of small differences sum to a big difference over many steps. Under Edgington's treatment, a series of reasonable moves, each of which carries a small risk, can mount over enough repetitions to make for an unreasonable move carrying a high risk.

9.2 Sorites: The Final Analysis

Viewed formally or informally, from the perspective of truth, logic, or meaning, the Sorites argument does not pose a serious problem for the account developed over the course of this thesis. This is confirmed by the answers that can be given to the questions posed at the outset of the chapter. (Q1) Sorites arguments are valid. (Q2) The argument is unsound. (Q2a) This is not because any single premise is simply false or entirely unreasonable. Each step carries a slight risk, a tiny chance that the classifier is not being used as it should be. This does not just mean a risk of speaking falsely. On the pragmatic view, being true or false is just one property words can have. Rather, the risk is of using words ineffectively, of failing to use words in line with communicative norms. (Q2b) This explains why we cannot simply pinpoint THE false premise. No premise is entirely false. Depending on the modelling of the tolerance conditionals, no tolerance conditional carries a greater risk/greater degree of falsity than any other. Furthermore, the risk of each conditional is very small.

It is for good reason that we usually ignore such tiny risks (§8.1.3). To not do so would put too great a strain on our cognitive load. However, this generally successful strategy of ignoring risks is exploited to ill effects by a well constructed Sorites. If we do take account of such risks, we can see how they mount.

Finally, (Q3), to know the meaning of a term (to be competent with its use) is, in part, on this account, not to know exactly where to draw a line. The competent thing to do in some cases is to choose NOT to choose to make a classification. Metalinguistic uncertainty over classification is not uncertainty about which use of a classifier is true (over how things are). It is uncertainty about how speakers use classifiers to some end.

Put formally, if one restricts the probability space to two possibilities: the production of one classification, and the production of its negation, then a speaker has the same probability of making the classification as making the negation. More realistically, a speaker would make neither the classification nor its negation because it is highly unlikely that either a classifier or its negation would prove successful for many purposes in borderline cases. Just imagine being faced with two almost identical looking swatches. Trying to refer by uttering 'The red one' (or 'The not-red one') is unlikely to prove effective. The competent thing to do is to choose not to choose either 'red' or 'not red'. The sensible thing to do is to choose words that will serve one's purpose, for example, 'The swatch on the left'. Being a competent speaker of a language necessitates being able to shift and adjust in what one says to serve the situation in which one finds oneself. Sorites arguments force a false dichotomy in restricting choices to one of two classifier judgements. So restricted, at some point, one must choose not to choose. If speakers do this at different stages in the series (or, indeed, if individual speakers show hesitance at different points on different occasions), this does not imply much about the meaning of the classifier in question. Rather, on the view proposed, this reflects how the very rough and ready practical abilities that we have (for balancing the reasons for judgements that our linguistic knowledge provides us with) do not always leave a sharp line in the sand.

10

Summary and Conclusions

In this, the final chapter, we will revisit five main topics that have been recurrent throughout: *Information and Communication*, *Learning*, *Meaning and Context*, *Vagueness*, and *Truth, Logic, and the Sorites*. For each of these, the main points that have been argued for will be summarised. Finally, some overall conclusions will be drawn.

10.1 Information and Communication

Grice's influential paper *Meaning* (Grice 1957) starts out with the idea that the sort of meaning that words or intentional actions have is different to the kind of meaning that natural signs carry. The same kind of distinction transposes into information theoretic talk. In our environments there are many signs which carry natural information. Natural signs cannot misrepresent. In contrast, non-natural (e.g. linguistic) signs can both represent and misrepresent, and so linguistic signs cannot always carry natural information. Representing and misrepresenting requires another explanation (for example, in terms of evolved functions or faulty mechanisms).

Chapters 2-4 could be read as a response to the kind of argument in the previous paragraph. In the above argument, representing is a euphemism for having truth-conditions (for saying that things are some specific way). That is why non-natural signs can also misrepresent (can say that things are as they are not). The arguments I provided against this conclusion were two-pronged:

First, I argued that there are severe problems with viewing words as being representational (in the sense of having truth-conditions), and that some ways of specifying semantic information were just ways of expressing the same problematic idea. That brings into doubt the premise that non-natural signs can both represent and misrepresent (on the truth-conditional understanding of 'representation').

Second, appealing to Millikan, I argued that natural information should be viewed as correlative not law-governed or truth-conditional. This allowed us to resist the conclusion that linguistic signs frequently fail to carry information at all.

There is, however, something that Grice's distinction captures. There is a distinction to be drawn between natural and non-natural signs. The difference, I suggested, lies in the reasons why the correlations they indicate persist. Linguistic signs, like natural signs, only indicate that states of affairs are probable. However, linguistic signs, unlike natural signs, can be used, copied, reused, and misused. It is unclear how one could copy or reuse a natural sign (a shadow that passes on the ground, the rings on a tree stump). But linguistic signs are easily replicable. 'The reproduction of signs, at least in part due to precedent' sums up Millikan's view of conventions that I adopted. The difference between natural and non-natural signs was that the correlations that they indicate are sustained for different reasons. In the natural case, laws might frequently contribute to why statical correlations persist. In the linguistic case, correlations persist because they are reproduced to bring about the same kinds of effects.

Different models of communication all agree that part of what goes on in communication is the encoding and decoding of information. All of the arguments surrounding information were also arguments about what is encoded (and decoded) by words. There is, of course, more to communication than encoding and decoding. However, describing what is encoded is a way of spelling out the basic framework upon which more complex communication can be built. On the view of semantic information I defended, questions about communicative acts became questions about how the information carried by words can be used to perform acts. This is because, for example, if words don't themselves say something true, an account of saying something true must appeal to how the information carried by words can be used.

10.2 Learning

Semantic learning came into the arguments both as a negative, and as a positive force. On the negative side, learning was used to argue that some accounts of meaning and information are not learnable. On the positive side, considerations about the kinds of situations we are exposed to as learners helped further support the idea that semantic information must be probabilistic in character. The main driving force behind these arguments was the bottleneck problem. We all have a limited amount of experience of language use in the learning phase of our development. How is it that we all learn to use terms to describe novel objects in novel situations, and yet do so with a large amount of stability (we all come to use terms in approximately similar ways).

What learners need is the ability not just to record instances of uses of words, but to form generalisations about how those words are used. I argued that probabilistic learning seems to be the only account with the right kind of features to form such generalisations because more declarative generalisations would produce unstable and inconsistent uses of terms both within an individual's linguistic behaviour and more widely across a linguistic community.

10.3 Meaning and Context

The account of semantic information meant that meanings could not be stated in terms of truth conditions. Instead, I argued for the view that we can characterise what words mean in terms of the information they carry and what that entitles us to believe. Meanings of words were therefore characterised as what it would be reasonable to believe, given that those words had been used.

If words carry information that is grounded in correlations, then what words encode is non-specific. But we frequently need to be more precise than the information our words, on their own, carry. Relatedly, we need an account of how non-specific words can be used to bring about specific ends and specific communicative goals. Chapters 6 and 7 addressed these problems in two ways (i) via the use of other words, and (ii) via the use of the situations in which we use them:

(i) The information that words carry can compose into more specific information. Combinations of information carried by words can lead to highly specific messages even from highly unspecific components. We saw cases of this in chapter 6. For example, adjectives select information carried by other words and modify this information to give more specific expectations about the situation being described.

(ii) Nonetheless, even the information carried by combinations of words needs to be further constrained in order to achieve specific ends. The semantic information carried by words was integrated with occasion specific information from resource situations. For example, the fact that interlocutors are attending to a restricted class of objects allows one to infer what specific object is being referred to even if the words themselves are non-specific. It is in that sense that words (and the information they carry) can be used by agents to achieve certain goals.

10.4 Vagueness

This account of the information that words carry led to a natural explanation of vagueness. With respect to borderline cases, vagueness falls naturally out of the account of meaning. If meanings are characterised in terms of reasonable expectations (where these are graded over numbers of possibilities), then we will have as much reason to classify some objects one way as we do another. I suggested that other aspects of vagueness could be captured by appealing to the non-specificity of words and how they are used. For example, a description can be vague because words are used in such a way that the information they carry and the information contained in the situation they describe do not combine to give specific enough reasonable expectations to the addressee. Take the case where there are two houses, one is light blue, and the other is dark blue. To try to refer to the light blue house just by using the words 'The blue house', would be to be vague in the sense that one is not being specific enough for one's purpose.

10.5 Truth, Logic, and the Sorites

Even if meaning is not identified with truth conditions, we can still use words to truly or falsely describe the world. The account of truth I have defended is pragmatic, or perhaps pragmatist. Truly describing the world is inextricably linked to the purposes at hand and the goals one needs to achieve. However, I argued that this does not imply a relativism about truth, nor is it anti-realist. Indeterminacy and borderline cases arise as a side-effect of the information words encode. As we have already noted, vague words can be used to achieve specific goals. Falsity arises from using words in such a way to make it reasonable for someone to expect the world to be a way that it is not. Saying something true can simply be a matter of making it reasonable for someone to expect that the world is the way it is. However, truth is pragmatic, not least because the amount of specificity required to say how the world is can vary.

When truth is tied, to some extent, to probability (reasonable expectations), we have an excellent reason for why a logic for borderline cases should be the same as a logic for uncertainty. I defended using Bayesian classical probability logic as a logic for borderline cases. Such a logic is equivalent to classical Boolean logic at the limit. Between the limits it does not suffer from the problems faced by other accounts such as fuzzy logic and supervaluationism.

At the outset of this thesis I suggested that the place for the Sorites paradox in an account of vagueness is as a test (albeit not a definitive one). The account of truth and meaning that I have defended does not lead to contradiction when faced with the Sorites. Informally, we can see the Sorites as exploiting our tendency to ignore small reasons for making linguistic classifications. That is why we find it compelling and also why we are led towards contradiction. Formally, using Edgington's solution to the paradox, the Sorites can be seen to be valid but unsound. It is unsound not because some single premise is plain false, but because small risks mount over repeated inferences.

10.6 Concluding Remark

To be learnable, words must contribute something that is pretty stable across contexts because what they carry must be capable of being estimated, given a limited experience of language use (albeit employing some extra learning and reasoning abilities). But equally, words must also be flexible enough to be able to stretch, in a principled way, to cover new cases (so that language can remain stable). Similarly, to be effective for communication, the information that words carry must be stable enough to be decodable, but flexible enough to be applicable to new and varied cases. I have argued that truth conditions, and information understood in terms of truth conditions, cannot satisfy these requirements. My replacement of truth conditions with probabilities, uncertainty, and reasonable belief will no doubt prove to face major challenges itself. I hope, however, that I have gone some way towards giving an explanation of how vagueness (as in borderline cases) can arise out of the pressures that our words face as learnable tools for communication. Just as importantly, I also hope to have shown that, when the meanings of words are viewed in a different way, what it is to be vague in the use of words can be understood a little more clearly.

References

Armstrong, D. M., 1989. *Universals: An Opinionated Introduction*. Westview Press.

Austin, J. L., 1950/1979. "Truth". In: Urmson, J. O., Warnock, G. J. (Eds.), *Philosophical Papers, Third Edition*. Originally in: *Symposium: Truth, Proceedings of the Aristotelian Society*, Supplementary Volume Vol. 24 (1950). Oxford University Press, Oxford, pp. 117–133.

Austin, J. L., 1961/1979. "Unfair to Facts". In: Urmson, J. O., Warnock, G. J. (Eds.), *Philosophical Papers, Third Edition* (First edition published 1961). Clarendon Press, pp. 154–174.

Austin, J. L., 1962. *Sense and Sensibilia*. Clarendon, Oxford.

Ayer, A. J., 1935. "The Criterion of Truth". *Analysis* 3, 28–31.

Bach, K., 1994. "Conversational Implicature". *Mind and Language* 9 (2), 124–162.

Bar-Hillel, Y., Carnap, R., 1952. "An Outline of a Theory of Semantic Information". MIT Research Laboratory of Electronics, *Technical Report* (247), 1–49.

Barr, D. J., 2004. "Establishing Conventional Communication Systems: Is Common Knowledge Necessary?". *Cognitive Science* (28), 937–962.

Barwise, J., 1989. *The Situation in Logic*. CSLI.

Barwise, J., Perry, J., 1983. *Situations and Attitudes*. MIT Press.

Braun, D., Sider, T., 2007. "Vague, So Untrue". *Noûs* Vol. 41 No. 2, 133–156.

Brighton, H., Kirby, S., 2001. "Meaning Space Structure Determines the Stability of Culturally Evolved Compositional Language". *Technical report*, Language Evolution and Computation Research Unit, Department of Theoretical and Applied Linguistics, The University of Edinburgh.

Cappelen, H., Lepore, E., 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Blackwell, Oxford.

Carnap, R., 1962. *Logical Foundations of Probability*. Chicago, University of Chicago Press.

Carpenter, B., 1997. *Type-Logical Semantics*. MIT Press.

Cooper, R., 1996. "The Role of Situations in Generalized Quantifiers". In: Lappin, S. (Ed.), *The Handbook of Contemporary Semantic Theory*. Blackwell, Cambridge, MA, pp. 65–86.

Cooper, R., 2005. "Austinian Truth, Attitudes and Type Theory". *Research on Language and Computation* Vol. 3, 333–362.

Cooper, R., 2012. "Type Theory and Semantics in Flux". In: Kempson, R., Fernando, T., Asher, N. (Eds.), *Philosophy of Linguistics*, Handbook of the Philosophy of Science. Elsevier, pp. 271–323.

Cooper, R., Dobnik, S., Lappin, S., Larsson, S., 2013. "Probabilistic Type Theory and Natural Language Semantics" ms, University of Gothenburg and King's College London.

Cresswell, M., J., 1974. "The Semantics of Degree". In: Partee, B. (Ed.), *Montague Grammar*. New York: Academic Press, pp. 261–292.

Davidson, D., 1965/1984. "Theories of Meaning and Learnable Languages". In: *Inquiries Into Truth And Interpretation Vol. 2*. Originally published in *Proceedings of the International Congress for Logic, Methodology, and Philosophy of Science pp. 3–17*. North-Holland, pp. 4–16.

Davidson, D., 1967. "Truth and Meaning". *Synthese* 17 (1), 304–323.

Davidson, D., 1969. "True to the Facts". *Journal of Philosophy* 66 (21), 748–764.

Davidson, D., 1973. "Radical Interpretation". *Dialectica* 27 (1), 314–328.

Davidson, D., 1974. "Belief and the Basis of Meaning". *Synthese* 27 (July-August), 309–323.

Davidson, D., 1977. "Reality without Reference". *Dialectica* Vol. 31 (3-4), 247–258.

Davidson, D., 1984. "Moods and Performances". In: *Inquiries Into Truth And Interpretation Vol. 2.* Oxford University Press, pp. 110–122.

Davies, A., 2011. *A Purpose for Context Sensitivity* PhD Thesis, King's College London.

Devlin, K., 2006. "Situation Theory and Situation Semantics". In: Gabbay, D., Woods, J. (Eds.), *Handbook of the History of Logic*, Vol. 7. Elsevier, pp. 601–664.

Dretske, F., 1981. *Knowledge and the Flow of Information.* MIT Press.

Dretske, F., 1988. *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

Dummett, M., 1959. "Truth". *Proceedings of the Aristotelian Society* 59 (1), 141–62.

Edgington, D., 1992. "Validity, Uncertainty and Vagueness". *Analysis* 52 (4), 193–204.

Edgington, D., 1997. "Vagueness by Degrees". In: Keefe, R., Smith, P. (Eds.), *Vagueness: A Reader.* MIT Press, Cambridge, MA, pp. 294–316.

Edgington, D., 2010. "Sorensen on Vagueness and Contradiction". In: Dietz, R., Moruzzi, S. (Eds.), *Cuts and Clouds–Vagueness, its Nature, and its Logic.* Oxford University Press, Oxford, pp. 91–106.

Fetzer, J., 2004. "Information: Does it Have To Be True?', *Minds and Machines* Vol. 14, 223–229.

Field, H., 1972. "Tarski's Theory of Truth". *Journal of Philosophy* 64 (13), 347–375.

Field, H., 1994. "Deflationist Views of Meaning and Content". *Mind* 103 (411), 249–285.

Fine, K., 1975. "Vagueness, Truth and Logic", *Synthese* Vol. 30, 265–300.

Floridi, L., 2011. *The Philosophy of Information*. Oxford University Press, Oxford.

Fodor, J. A., LePore, E., 1991. "Why Meaning (Probably) Isn't Conceptual Role". *Mind and Language* 6 (4), 328–43.

Fodor, J. A., Lepore, E., 1996. "What Cannot Be Evaluated Cannot Be Evaluated and It Cannot Be Supervalued Either". *Journal of Philosophy* 93 (10), 516–535.

Frazee, J., Beaver, D., 2010. "Vagueness is Rational Under Uncertainty". In: Aloni, M., Bastiaanse, H., de Jager, T., Schulz, K. (Eds.), *Logic, Language and Meaning*: 17th Amsterdam Colloquium, Amsterdam, The Netherlands, December 16-18, 2009, Revised Selected Papers, Lecture Notes in Artificial Intelligence. Springer, pp. 153–162.

Frege, G., 1892/1948. "Sense and Reference". Originally (1892) 'Über Sinn und Bedeutung', in Zeitschrift für Philosophie und philosophische Kritik, 100: 25-50. *The Philosophical Review* Vol. 57, 209–230.

Frege, G., 1918/1956. "The Thought: A Logical Enquiry". *Mind* Vol. 259, 289–311.

Geach, P. T., 1962. *Reference and Generality*. Ithaca, N.Y.,Cornell University Press.

Goguen, J. A., 1969. "The Logic of Inexact Concepts". *Synthese* 19 (3-4), 325–373.

Graff Fara, D., 2000. "Shifting Sands: An Interest-Relative Theory of Vagueness". *Philosophical Topics* 28 (1), pp. 45–81.

Gregoromichelaki, E., Kempson, R., Purver, M., Mills, G., Cann, R., Meyer-Viol, W., Healey, P., 2011. "Incrementality and Intention-Recognition in Utterance Processing". *Dialogue and Discourse* 2 (1), 199–232.

Grice, P., 1957. "Meaning". *The Philosophical Review* Vol. 66 (3), 377–388.

Grice, P., 1989a. "Logic and Conversation". In: *Studies in the Way of Words*. Harvard, London, pp. 22–40.

Grice, P., 1989b. *Studies in the Way of Words*. Harvard, London.

Griffiths, T., Kemp, C., Tenenbaum, J., 2008. "Bayesian Models of Cognition". In: Sun, R. (Ed.), *The Cambridge Handbook of Computational Psychology*. Cambridge University Press.

Groenendijk, J., Stokhof, M., 1991. “Dynamic Predicate Logic”. *Linguistics and Philosophy* 14 (1), 39–100.

Horwich, P., 1990. *Truth*. Blackwell, Oxford.

Horwich, P., 1998. *Meaning*. Oxford University Press.

Kamp, H., 1975. “Two Theories about Adjectives”. In: Keenan, E. (Ed.), *Formal Semantics of Natural Language*. Cambridge University Press, pp. 123–155.

Kamp, H., 1981. “The Paradox of the Heap”. In: *Aspects of Philosophical Logic*. Dordrecht. Reidel, pp. 225–227.

Kamp, H., Reyle, U., 1993. *From Discourse to Logic*. Kluwer, Dordrecht.

Kaplan, D., 1989. “Demonstratives”. In: Almog, Perry, Wettstein (Eds.), *Themes from Kaplan*. Oxford University Press, Oxford, pp. 581–563.

Kearns, S., Magidor, O., 2008. “Epistemicism About Vagueness and Meta-Linguistic Safety”. *Philosophical Perspectives* 22 (1), 277–304.

Keefe, R., 2000. *Theories of Vagueness*. Cambridge University Press.

Kemp, C., Tenenbaum, J., 2008. “The Discovery of Structural Form”. *Proceeding of the National Academy of Sciences* 105 (31), 10687–10692.

Kennedy, C., 1997. *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. PhD Thesis. University of California, Santa Cruz.

Kennedy, C., 2007. “Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives”. *Linguistics and Philosophy* 30 (1), 1–45.

Keysar, B., Barr, D. J., Horton, W. S., 1998. “The Egocentric Basis of Language Use: Insights from a Processing Approach”. *Current Directions in Psychological Sciences* 7, 46–50.

Kirby, S., 2007. “The Evolution of Meaning-Space Structure through Iterated Learning”. In: Lyon, C., Nehaniv, C., Cangelosi, A. (Eds.), *Emergence of Communication and Language*. Springer, Verlag, London, pp. 253–268.

Kirby, S., Hurford, J., 2002. “The Emergence of Linguistic Structure: An overview of the Iterated Learning Model”. In: Cangelosi, A., Parisi, D.

(Eds.), *Simulating the Evolution of Language*. Springer, Verlag, London, pp. 121–148.

Kolmogorov, A., 1950. *Foundations of Probability*. Chelsea Publishing, New York.

Lasersohn, P., 1999. Pragmatic Halos. Language 75 (3), 522–551.

Lassiter, D., 2011. “Vagueness as Probabilistic Linguistic Knowledge”. In: Nouwen, R., Sauerland, U., Schmitz, H., van Rooij, R. (Eds.), *Vagueness in Communication*. Springer.

Lewis, D., 1969. *Convention: A Philosophical Study*. Harvard University Press.

Lewis, D., 1970. “General Semantics”. *Synthese* 22 (1-2), 18–67.

Lewis, D., 1993. “Many, but Almost One”. In: *Ontology, Causality, and Mind: Essays on the Philosophy of D. M. Armstrong*. Cambridge University Press, pp. 23–38.

MacFarlane, J., 2009a. “Nonindexical Contextualism”, *Synthese* Vol. 166 No. 2, 231–250.

MacFarlane, J., 2009b. “Fuzzy Epistemicism”. In: Dietz, R., Moruzzi, S. (Eds.), *Cuts and Clouds*. Oxford University Press, pp. 438–463.

McDowell, J., 1976. “Truth-Conditions, Bivalence, and Verification”. In: *Truth and Meaning*. Clarendon Press.

Millikan, R. G., 1984. *Language, Thought and Other Biological Categories*. MIT Press.

Millikan, R. G., 2000. *On Clear and Confused Ideas: An Essay About Substance Concepts*. Cambridge University Press.

Millikan, R. G., 2004. *Varieties of Meaning*: The 2002 Jean Nicod Lectures. MIT Press.

Millikan, R. G., 2005a. “Language Conventions Made Simple”. In: *Language: A Biological Model*. Oxford: Clarendon Press.

Millikan, R. G., 2005b. *Language: A Biological Model*. Oxford: Clarendon Press.

Niiniluoto, I., 1977. "On the Truthlikeness of Generalizations". In: Butts, Hintikka (Eds.), *Basic Problems in Methodology and Linguistics*. Dortrecht:Reidel, pp. 121–147.

Niiniluoto, I., 1978. "Truthlikeness: Comments on Recent Discussion". *Synthese* 38 (2), 281–329.

Peirce, C. S., 1877. "Illustrations of the Logic of Science: First Paper – The Fixation of Belief". *Popular Science Monthly* 12, 1–14.

Popper, K., 1963. *Conjectures and Refutations*. London: Routledge.

Predelli, S., 2005. "Painted Leaves, Context, and Semantic Analysis". *Linguistics and Philosophy* 28 (3), 351–374.

Putnam, H., 1975. "The Meaning of 'Meaning'". *Minnesota Studies in the Philosophy of Science* 7, 131–193.

Putnam, H., 1983. "Vagueness and Alternative Logic". *Erkenntnis* 19 (1-3), 297–314.

Putnam, H., 1985. A Quick Read is a Wrong Wright. Analysis 45 (4), 203 –.

Quine, W. V., 1953. *From a Logical Point of View*. Harvard University Press.

Quine, W. V., 1960. *Word and Object*. The MIT Press.

Quine, W. V., 1968. "Ontological Relativity". *Journal of Philosophy* 65 (7), 185–212.

Quine, W. V., 1970/1986. *Philosophy of Logic*. Harvard University Press.

Raffman, D., 1994. "Vagueness Without Paradox". *Philosophical Review* 103 (1), 41–74.

Raffman, D., 1996. "Vagueness and Context-Relativity". *Philosophical Studies* Vol. 81, 175–192.

Ramsey, F. P., 1927. "Facts and Propositions". *Proceedings of the Aristotelian Society (Supplementary)* 7 (1), 153–170.

Rayo, A., 2011. "A Plea for Semantic Localism". *Noûs*, 1–35.

Richard, M., 2000. "On an Argument of Williamson's". *Analysis* 60 (266), 213–8211.

Sainsbury, M., 1997. "Concepts Without Boundaries" in. In: Keefe, R., Smith, P. (Eds.), *Vagueness: A Reader*. MIT Press, Cambridge, MA, pp. 251–264.

Schiffer, S. R., 2003. *The Things We Mean*. Oxford University Press.

Schwartz, S. P., Throop, W., 1991. "Intuitionism and Vagueness". *Erkenntnis* 34 (3), 347–356.

Searle, J. R., 1968. "Austin on Locutionary and Illocutionary Acts". *Philosophical Review* 77 (4), 405–424.

Shannon, C., 1948. "A Mathematical Theory of Communication". *Bell System Technical Journal* 27, 379–423.

Skyrms, B., 2010. *Signals: Evolution, Learning, and Information*. OUP Oxford.

Smith, N. J. J., 2008. *Vagueness and Degrees of Truth*. Oxford University Press.

Soames, S., 1999. *Understanding Truth*. Oxford University Press.

Sorensen, R., 1988. *Blindspots*. Clarendon Press, Oxford.

Sorensen, R., 2001. *Vagueness and Contradiction*. Clarendon Press, Oxford.

Sperber, D., Wilson, D., 1986. *Relevance, Communication and Cognition*. Blackwell, Oxford.

Strawson, P. F., 1949. "Truth". *Analysis* 9 (6), 83–97.

Strawson, P. F., 1950. "Truth". *Symposium: Truth, Proceedings of the Aristotelian Society, Supplementary Volume* Vol. 24, 129–156.

Tappenden, J., 1993. "The Liar and Sorites Paradoxes: Toward a Unified Treatment". *Journal of Philosophy* 60 (11), 551–577.

Tarski, A., 1944. "The Semantic Conception of Truth: And the Foundations of Semantics". *Philosophy and Phenomenological Research* 4 (3), 341–376.

Tarski, A., 1956. "The Concept of Truth in Formalized Languages". In: *Logic, Semantics, Metamathematics*. Oxford University Press, pp. 152–278.

Tenenbaum, J., Kemp, C., Griffiths, T., Goodman, N., 2011. "How to Grow a Mind: Statistics, Structure, and Abstraction". *Science* 331, 1279–1285.

Travis, C., 1978. "Meaning Versus Truth". *Dialogue* 17 (03), 401–430.

Travis, C., 2008a. "Annals of Analysis". In: *Occasion-Sensitivity: Selected Essays*. Oxford University Press, Oxford, pp. pp. 65–93.

Travis, C., 2008b. "Meaning's Role in Truth". In: *Occasion-Sensitivity: Selected Essays*. Oxford University Press, Oxford, pp. pp. 94–108.

Travis, C., 2008c. "On What is Strictly Speaking True". In: *Occasion-Sensitivity: Selected Essays*. Oxford University Press, Oxford, pp. pp. 19–64.

Travis, C., 2008d. "Pragmatics". In: *Occasion-Sensitivity: Selected Essays*. Oxford University Press, Oxford, pp. pp. 109–129.

Travis, C., 2008e. "Vagueness, Observation and Sorites". In: *Occasion-Sensitivity: Selected Essays*. Oxford University Press, Oxford, pp. pp. 206–226.

Tversky, A., Kahneman, D., 1974. "Judgment Under Uncertainty: Heuristics and Biases". *Science* 185, 1124–1131.

Unger, P., 1980. "The Problem of the Many". *Midwest Studies in Philosophy* 5 (1), 411–468.

van Eijck, J., Lappin, S., 2012. "Probabilistic Semantics for Natural Language", unpublished ms, CWI Amsterdam and King's College London.

Weaver, W., 1979. "The Mathematics of Communcation". In: Mortensen, C. D. (Ed.), *Basic Readings in Communication Theory*. Harper and Row, pp. 27–38 Originally published in *Scientific American* 181 (1).

Williams, J. R. G., 2011. "Degree Supervaluational Logic". *The Review of Symbolic Logic* 4, 130–149.

Williamson, T., 1994. *Vagueness*. Routledge, Abingdon.

Wittgenstein, L., 1953/2009. *Philosophical Investigations* Revised 4th Edition, Hacker and Schulte (eds.). Blackwell, Oxford.

Wright, C., 1975. "On the Coherence of Vague Predicates". *Synthese* 30 (3-4), 325–365.

Zadeh, L., 1965. "Fuzzy Sets". *Information and Control* 19 (8), 328–353.